

Dynamic and Computational Models of Recurrent Networks in the Brain

Thesis submitted for the degree
"Doctor of Philosophy"

Oren Shriki

Submitted to the Senate of the Hebrew University in the year 2003

This work was carried out under the supervision of

Prof. Haim Sompolinsky

Dedicated in memory of my beloved father Motti Shriki who taught me to think differently and not to compromise

Acknowledgements

"When we will discover the laws underlying natural computation, we will finally understand the nature of computation itself."

John von Neumann

The effort that was made in this work to take a small step forward in our understanding of information processing in the brain would not have been accomplished without the help of many people. Some helped with understanding 'computation in nature,' some with 'the nature of computation' and some with both.

First, I am greatly indebted to my advisor, Prof. Haim Sompolinsky, for the many years of guidance and teaching he dedicated to our research. We spent many hours of discussion, brainstorming, mathematical calculations and programming together. He also applied his considerable skills to make the presentation of the works as smooth and coherent as possible while maintaining scientific accuracy.

Special thanks go to Prof. David Hansel who taught me the skill of numerical simulations of conductance-based neurons, helped and gave his advice in many discussions that we had throughout the years and was a coauthor of the paper that appears in Chapter 2.

Additional thanks to Dr. Dan Lee who coauthored the paper that appears in Chapter 3. I learned important lessons from him about information theoretic models and their applications.

I would like to extend warmest thanks to my friends in the neurophysics lab: Maoz Shamir, Jonathan Loewenstein, Josh Goldberg and Uri Rokni. They were always there to share and discuss ideas, give their viable help in solving problems and provide new insights into the research. The warm and stimulating atmosphere in our group had a crucial impact on this work.

I am particularly grateful to Alisa Shadmi who was the secretary of the Center for Neural Computation during these years. Her encouragement and emotional support along the way (even after she left the center) were important.

Further thanks to Esther Singer who skillfully went over the English in this manuscript.

To my beloved parents who always stood behind me. I am grateful to my mother for supporting and helping and often taking care of my children, allowing me to dedicate more time to the work. My father who was also involved and interested in my work, did not live to see this manuscript but I promised him to finish the job and I am sure he knows that I kept my promise.

To my beloved wife Galit who shared with me the wonderful moments of new insights and enthusiasm as well as the tough ones. She was often 'forced' to hear lectures about my progress and I am grateful to her for listening and giving me inspiration.

Finally, to the young 'professors' I have at home, Yaara and Ori, who fill life with joy.

There were, obviously, many more people than the aforementioned that contributed to this work. I thank you all.

Abstract

This thesis focuses on two fundamental questions in modeling of recurrent neuronal networks in the brain: (1) How can the dynamics of large cortical neuronal networks be described by means of simplified models and (2) What is the functional role of recurrent interactions among cortical neurons in sensory processing. Chapter 2 presents the framework we have developed to answer the first question. Chapters 3 and 4 concern the second question.

When modeling neuronal systems, a major issue is the relevant level of description of the underlying dynamics. Theoretical models of the collective behavior of large neuronal systems tend to fall into two main categories. One approach is to construct detailed biophysical models, which are then studied using numerical simulations. Models of this type describe the dynamics of conductances in nerve cells and are termed *conductance-based* models. The neuronal activity in these models is characterized by the behavior of the voltage as a function of time. The second approach is to consider simplified models of the neuronal behavior that allow for analytical investigation. In these models, the dynamic variable that characterizes the neuronal activity is the firing rate of the neuron, i.e., the number of electrical pulses (action potentials) in a unit of time.

Rate models provide powerful tools for investigating the principles that underlie the cooperative functions of large neuronal systems. However, the biophysical interpretations of these models have been ambiguous, limiting their applicability to modeling real neuronal systems and their possible experimental validation. In Chapter 2, we show that conductance-based models of large cortical neuronal networks can be described by simplified rate models, provided that the network state does not possess a high degree of synchrony. We derive circuit equations for the firing rates, which have the familiar form of neural-network like rate models. Since the model is derived explicitly from a general conductance-based model, it provides a clear biophysical interpretation for the parameters of the rate model. We express this in the form of a precise mapping between the familiar rate model parameters and those of the conductance-based model. The model allows for a quantitative prediction of firing rates and synaptic conductances in conductance-based networks by solving relatively simple equations, and without resorting to extensive numerical simulations.

The basic idea of the approach is that in a large asynchronous network the changes in the input current to the cell and in the cell's conductance are approximately constant in time. In this sense, the effect of the network on the firing rate of a neuron can be reduced to a problem at the single neuron level. The issue is how to characterize the firing rate response of the neuron to constant external currents and to constant changes in its input conductance. We use a simple phenomenological model to describe the firing rate response properties of isolated cells. The main assumption is that increasing the cell's input conductance has a primarily subtractive effect on its frequency-current response curve: the current threshold increases, while the shape of the curve does not change. Another simplifying assumption is that the firing rate of a single cortical cell increases semi-linearly with the amplitude of a constant injected current. Both assumptions are consistent with experimental data on the frequency-current curves of cortical cells. To compare the predictions of the rate model to results from numerical simulations of conductance-based networks, we devised a single compartment Hodgkin-Huxley model with transient potassium A-current. Consistent with the above assumptions, this neuron has a semi-linear f-I curve and changing its input conductance has a subtractive effect.

As a simple demonstration of the rate model, we consider the case of a large homogeneous network with all-to-all excitatory interactions. It is shown that the firing rate of the neurons predicted by the rate model is highly congruent with simulation results in a broad range of synaptic conductances.

The approach is also applied to the study of a network model of a hypercolumn in the primary visual cortex. The network consists of cells that respond to bars with different orientations from the same patch of the visual field. The pattern of connections between cells that respond to different orientations has a "Mexican hat" shape. When the input to the network is uniform over all orientations the network presents two different behaviors depending on the spatial modulation of the recurrent interactions. When the modulation is small the network converges to a homogeneous state, i.e, all the neurons have the same level of activity. For strong enough modulation, the network develops a profile of activity that peaks at a certain orientation, depending on the initial activities of the neurons and on the input noise. This kind of behavior is termed *spontaneous symmetry breaking*. We show that the conditions for the stabilization of the homogeneous state in the conductance-

based network can be correctly predicted from the analytical solution of the corresponding rate model. Furthermore, under conditions in which the external input is non-uniform, the rate model provides accurate predictions for the profile of activity of the network and for the tuning curve of the synaptic conductances.

The original mapping of conductance-based networks onto rate models was developed for conditions in which the firing rates of the neurons are constant in time. We also explore extensions of the rate model to the dynamic domain, by studying the firing rate response of our conductance-based neuron to time dependent noisy inputs. We show that the dynamics of this response can be approximated using a time dependent second-order differential equation. This phenomenological single-cell rate model is used to generalize our network model to describe the response of conductance-based networks to time dependent inputs.

The other central issue of this thesis concerns with the computational significance of recurrent interactions, in particular in the context of sensory processing. Information from our senses is traditionally thought to be processed in a sequence of consecutive stages. At the macroscopic level these stage are associated with different brain areas, and at the microscopic level they are attributed to different layers of neurons. This idea has inspired models of sensory processing that are based on feedforward neural networks. However, a large body of anatomical and physiological findings indicate that in almost every part of the nervous system the dynamics also involve feedback. For instance, a large amount of the synapses that a cortical neuron receives are of cortical origin, most of them from nearby cells. This suggests that recurrent processing may have an important computational role.

From a computational point of view, a plausible assumption about the organization of the nervous system is that it develops in such a way as to optimize the transmission and representation of information. In Chapter 3 we use this approach to study the role of recurrent interactions in information representation. The network model consists of a layer of input units and a layer of nonlinear output units. There are feedforward connections from the input to the output layer and recurrent connections among the neurons of the output layer. The work focuses on the case where the number of output units is greater than the number of input units or equal to it. When an input is presented to the network it is processed by the recurrent dynamics and the resulting pattern of

activity is considered as the representation of that input. To quantify the quality of this representation we use the concept of *mutual information* from Shannon's Information theory. Maximizing the mutual information between the input and output with respect to both the feedforward connections as well as the recurrent interactions results in simple learning rules for both sets of parameters. The conventional independent components (ICA) learning algorithm can be recovered as a special case when there is an equal number of output units and no recurrent connections. We demonstrate these new learning rules on a simple two-dimensional input example. In this example, the optimal recurrent connections perform *gain control* on the inputs, i.e they prevent the neuronal activities from reaching saturation.

Chapter 4 applies the approach developed in Chapter 3 to a simplified network model of a hypercolumn in the primary visual cortex. We use the principle of maximum mutual information to evaluate the optimal pattern of recurrent interactions in such a network and its dependence on the statistics of the external inputs. When the characteristic contrast of the inputs to the network is small, the optimal profile of the recurrent interactions has a "Mexican hat" shape, with a modulation amplitude close to the symmetry breaking phase transition. This pattern of interactions greatly amplifies the response profile that is generated by the feedforward input. When the contrast is high the recurrent connections are inhibitory at the center and excitatory away from it. This pattern of connections reduces the saturation of the network. In both cases, the patterns of recurrent interactions allow the neurons to use their full dynamic range for information representation. The results have important implications for the computational role of intracortical recurrent excitation and inhibition in shaping the tuning properties of orientation selective neurons in cortex.

Contents

1	Introduction	1
1.1	Organization of Connectivity in the Brain - Feedforward, Recurrent and Feedback Connections	2
1.2	Mathematical Models of Neuronal Networks	6
1.2.1	Conductance-Based Models	6
1.2.2	Rate Models	12
1.3	Optimization Principles in Modeling of Neuronal Networks	13
1.4	Information Theory	14
1.5	Undercomplete, Complete and Overcomplete Representations	18
1.6	Independent Component Analysis	18
1.7	Orientation Selectivity in Primary Visual Cortex	20
1.7.1	The visual pathway and receptive fields	20
1.7.2	Orientation selective cells	21
1.7.3	Theories of orientation selectivity	23
2	Rate Models for Conductance Based Cortical Neuronal Networks	28
2.1	Abstract	28
2.2	Introduction	29
2.3	Models and Methods	30
2.3.1	Dynamics of a single-compartment cell	30
2.3.2	Network dynamics	33
2.3.3	Model of a hypercolumn in primary visual cortex	35

2.3.4	Numerical integration and analysis of spike responses	36
2.4	Rate Equations for General Asynchronous Neuronal Networks	37
2.5	Response of an Excitatory Population to a Time-independent Input	39
2.6	A Model of a Hypercolumn in Primary Visual Cortex	40
2.6.1	Emergence of a ring attractor	42
2.6.2	Tuning of firing rates and synaptic conductances	45
2.7	Rate Response of a Single Neuron to a Time-dependent Input	49
2.8	Response of a Neuronal Population to a Time Periodic Input	54
2.9	Discussion	58
2.10	Appendix A: Model Neuron	62
2.11	Appendix B: Model of a hypercolumn in V_1 : the mean-field equations	62
2.12	Appendix C: Rate dynamics in a neuronal population with uniform connectivity	64
3	An Information Maximization Approach to Overcomplete and Recurrent Representations	68
3.1	Abstract	68
3.2	Introduction	69
3.3	Information Maximization	69
3.4	Learning rules	71
3.5	Examples	72
3.5.1	Feedforward weights	72
3.5.2	Recurrent interactions	73
3.6	Discussion	74
3.7	Appendix 1: Relationship between input and output distributions	74
3.8	Appendix 2: Derivation of the learning rules	75
3.8.1	Feedforward weights	76
3.8.2	Recurrent interactions	76
4	The Role of Recurrent Interactions in Orientation Selectivity: an Information Maximization Approach	79

4.1	Abstract	79
4.2	Introduction	80
4.3	Network Structure	80
4.4	The Cost Function	81
4.5	The Low Contrast Case	82
4.5.1	Analytical derivation of the optimal pattern of interactions	83
4.5.2	Numerical simulations of the learning process	86
4.6	The High Contrast Case	87
4.7	Discussion	91
5	Discussion and conclusions	94
5.1	Rate Models for Conductance-based Networks	94
5.2	Optimizing Recurrent Network Architectures	96
5.3	The Role of Recurrent Interactions in Orientation Selectivity: an Information Maximization Approach	98
5.4	A First-principle Computational Approach to Quantitative Biophysical Predictions .	99

Chapter 1

Introduction

“Everything should be made as simple as possible, but no simpler”

Albert Einstein

The structure of this dissertation is as follows. Chapter 2 deals with the general subject of simplified models for neuronal networks, and when and how they are justified from a microscopic biological point of view. It introduces a framework for bridging between simplified rate models and detailed conductance-based models. This chapter appeared in the August 2003 volume of *Neural Computation*. Chapter 3 deals with the representation of information by networks of simplified neurons, and uses an information theoretic approach to derive general learning algorithms for the feedforward and recurrent interactions. It was published in the Proceedings of *Advances in Neural Information Processing Systems* in 2001. Chapters 2 and 3 can be read independently. However, the combination of both provides tools for producing quantitative biophysical predictions from a first-principle computational approach. Chapter 4 applies the theory developed in Chapter 3 to the issue of orientation selectivity in the primary visual cortex. It concentrates on the role of recurrent cortical interactions in the mechanism of orientation selectivity. In order to keep a uniform layout for this dissertation, the published papers were reproduced here from the original manuscripts. The text and figures are the same as in the original papers.

This introductory chapter is organized as follows. The first section presents a brief overview of the different types of connectivity patterns that are used for describing neuronal networks.

The second section deals with mathematical descriptions of neuronal networks. It provides the background for Chapter 2. The next four sections provide the background for Chapter 3. They focus on information representation in neuronal systems and on applications of *Information theory* to modeling network parameters. The last section introduces anatomical and physiological data related to early visual processing in the brain, and concerns the issue of orientation selectivity in the primary visual cortex. It provides additional background needed for Chapter 4.

When writing a Ph.D. thesis, a major question is “who is going to read it?”. My hope is that the contents of this thesis would help professional researchers as well as students. The level of the text is aimed mainly at a typical graduate student who has basic training in computational neuroscience. The above quote from Einstein probably refers to scientific theories, which should be elegant and simple but not trivial. The attempt here is to adopt this principle in the context of explaining scientific ideas as well.

1.1 Organization of Connectivity in the Brain - Feedforward, Recurrent and Feedback Connections

The classical view treats the processing of sensory information in the brain as a series of stages, performed one after the other. These stages are attributed to different brain areas that are connected in an anatomically sequential manner. For instance, in the main pathway of visual processing, information is transmitted from the retina to the lateral geniculate nucleus (LGN), then to the primary visual cortex (V1) and from there to two main sequences of cortical areas, termed the dorsal and ventral streams (see Fig. 1.1).

Historically, the sequential view of information processing in the brain can be traced back to theories developed during the beginning of the first millennium A.D [1]. Influenced by Greek scholars such as Galen and Plato, Christian scholars attributed all intellectual and mental abilities to the brain ventricles. The theory recognized three stages of processing; namely, acquisition of sensory information, judgment and sorting of this information, and finally storage of memories. The first ventricle, the anterior ventricle, was associated with sensory perception, the middle (today

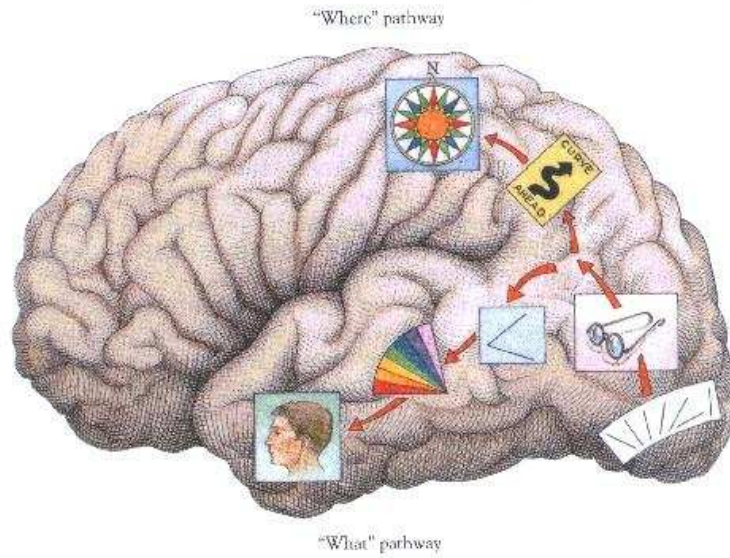
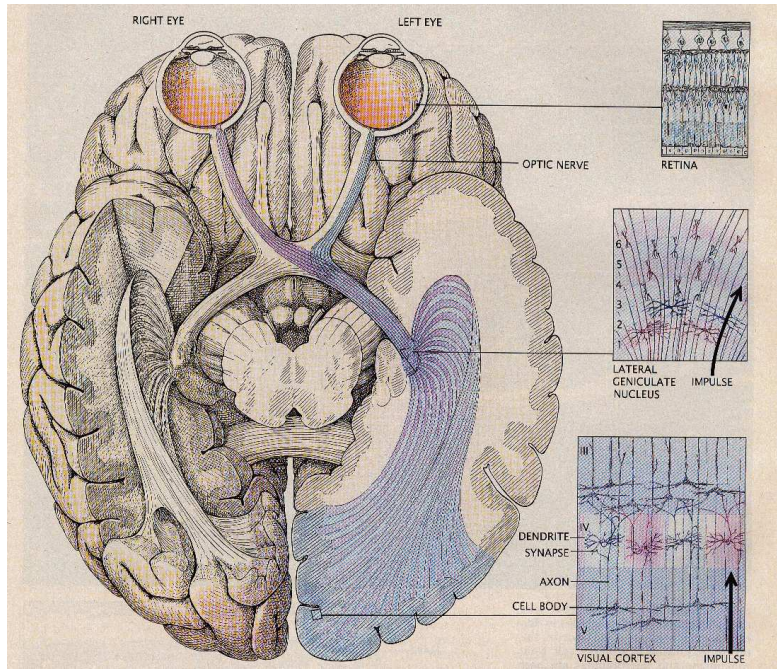


Figure 1.1: A. The visual pathway from the retina to the primary visual cortex. B. The dorsal and ventral streams of visual processing.

referred to as lateral) ventricles were associated with rational analysis, and the posterior ventricle with memory storage.

The idea that information is processed in a series of stages that are associated with consecutive brain areas led to the assumption that even at the microscopic level the processing of information is accomplished in a feedforward manner. This feedforward approach maintains that computation is performed by a layered network in which consecutive computational stages are carried out by consecutive layers of neurons connected by feedforward connections alone (see Fig. 1.2A). It should be noted that a given brain area may contain several processing layers, but the underlying architecture is still assumed to be feedforward.

Feedforward neural network models, inspired by the feedforward view of information processing, have been able to account for many experimental findings, such as the spatial properties of simple cells in the primary visual cortex (see Section 1.7).

However, in almost every part of the nervous system the dynamics also involve feedback. Formally, *Feedback* is said to exist in a dynamic system whenever the output of an element in the system influences in part the input applied to that particular element, thereby giving rise to one or more closed paths for the transmission of signals around the system [2]. Neural networks that contain feedback loops are termed *recurrent neural networks*.

In a feedforward network, the computation is carried out by the transformation from one layer to another. Normally, the input is introduced to the first layer and the output is read-out from the last layer. In contrast, in a recurrent network the computation is performed by the intrinsic dynamics of the network. The input is represented by the initial conditions and the steady-state response of the network represents the output.

In general, the addition of even one feedback loop to a feedforward network turns the network into a recurrent one. Nevertheless, in many instances it is still useful to picture the recurrent network as though it was organized in a layered structure (see Fig. 1.2B). As in a simple feedforward network, the notions of 'input' and 'output' layers are maintained, but there is a flow of signals inside a layer and back from a 'high level' layer to a 'low level' layer as well. The term *recurrent connections*, in this context, refers to interactions between neurons in the same processing layer. The term *feedback connections* refers to influences that go from a 'high level' layer to a 'low level'

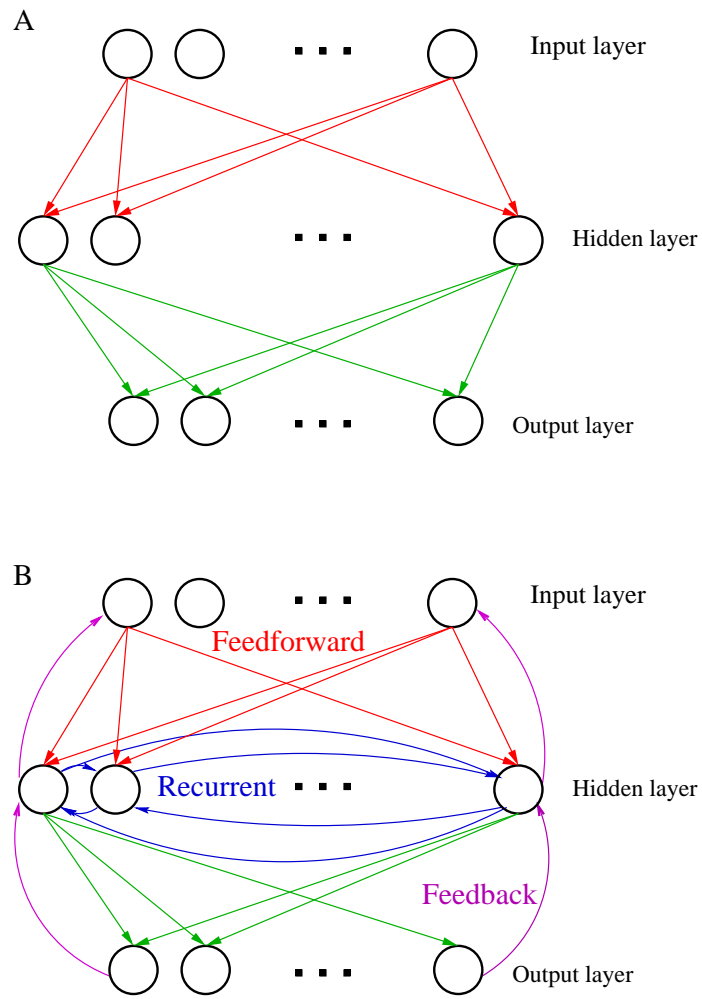


Figure 1.2: Networks architectures. A. Feedforward architecture B. A layered network with recurrent and feedback connections

layer.

It is important to note that in terms of the network dynamics the real difference is between feedforward processing versus recurrent processing. The above distinction between *recurrent connections* and *feedback connections* is only a matter of terminology when considering layered networks. This terminology is useful in the context of information processing in the brain. For instance, we can talk about feedforward connections from the LGN to V1, about recurrent connections inside the LGN or inside V1, and about feedback connections from V1 to the LGN or from the LGN to the retina.

1.2 Mathematical Models of Neuronal Networks

There are a wide range of approaches to mathematical modeling of neuronal networks. On one extreme are the detailed biophysical models which attempt to be loyal to the underlying biology. On the other extreme are the simplified models, often used to describe the collective behavior of large neuronal networks. Despite the usefulness of the simplified models, it is not always clear when they provide a relevant description of the biological system at hand, and what meaning should be assigned to the quantities and parameters used in them. Chapter 2 deals with this question. In this section we briefly describe the main features of conductance-based models, and of a popular class of simplified models called 'rate models'.

1.2.1 Conductance-Based Models

A large body of experimental evidence supports the view that on short time scales, information in the nervous system is represented by the electrical potential across the membrane of nerve cells. The dynamics of the membrane potential can be described by modeling the neuron as an electrical circuit. In general, the voltage of the neuron depends on both the time and the spatial location along the cell. This requires a description in terms of partial differential equations. Here we neglect the detailed morphology of nerve cells and focus on the so-called *point neuron model* which treats the neuron as a single isopotential compartment. The description of the neuron will thus be in terms of ordinary differential equations. For a detailed review of the subject, see [3, 4].

The membrane is composed of two layers of phospholipid molecules, that form together a

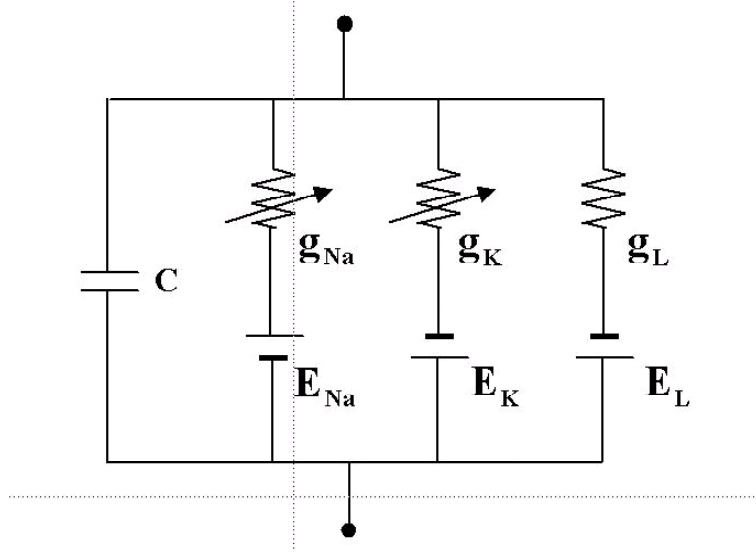


Figure 1.3: Equivalent electrical circuit of a membrane with sodium, potassium and leak currents.

thin insulating layer, separating the internal and external solutions. This structure acts like a *capacitance*. The *capacitive current* of the membrane is given by

$$I_C = C \frac{dV}{dt} \tag{1.1}$$

where C is the membrane capacitance and V is the membrane potential. This simply means that the capacitance limits the response rate of the membrane potential to changes in the current.

Proteins that cross the membrane allow for ions to move from one side to the other. These *ion channels* play the role of *conductances*. Specific types of ion channels are associated with specific ions, thus there is a different conductance for each ion type. For instance, sodium conductance is denoted by g_{Na} and potassium conductance is denoted by g_K .

The flow of a certain ion type through the membrane is influenced by the concentration gradient of this ion. In terms of the electrical circuit, each conductance can be considered as being serially connected to an *ionic battery*. For instance, the potentials of the batteries associated with sodium and potassium are denoted by E_{Na} and E_K , correspondingly. When the membrane potential crosses the potential of the battery of a certain ion, the current of this ion changes sign (say, from an inward to an outward current). For this reason, the voltage of the battery for a specific ion is also called the *reversal potential* of this ion. In general, these conductances are dynamical variables and are

not fixed with time. The fixed or passive component of the membrane conductance is usually called the *leak conductance* and is denoted by g_L . The reversal potential of the leak current is denoted by E_L . The equations for the sodium, potassium and leak currents are

$$I_{\text{Na}}(t) = g_{\text{Na}}(V(t), t)(V(t) - E_{\text{Na}}) \quad (1.2)$$

$$I_{\text{K}}(t) = g_{\text{K}}(V(t), t)(V(t) - E_{\text{K}}) \quad (1.3)$$

$$I_L(t) = g_L(V(t) - E_L) \quad (1.4)$$

Fig. (1.3) depicts the equivalent electrical circuit of a membrane with sodium, potassium and leak currents. To write the differential equation that governs the voltage dynamics, we use Kirchhoff's law of current conservation. The law states that *the sum of all currents flowing into and out of any particular node must equal zero*. We thus get

$$C \frac{dV}{dt} = -I_L - I_{\text{Na}} - I_{\text{K}} \quad (1.5)$$

$$= -g_L(V - E_L) - g_{\text{Na}}(V - E_{\text{Na}}) - g_{\text{K}}(V - E_{\text{K}}) \quad (1.6)$$

If all the conductances were constant in time, the voltage dynamics would be like in a simple RC circuit, and the effect of a constant current would simply be to charge the membrane. To produce the spiking behavior seen in experiments, Fig. (1.4), a dynamic model for the conductances has to be specified as well. We now introduce the standard kinetic model for conductance dynamics introduced by Hodgkin and Huxley in 1952.

The model expresses each of the dynamical ionic conductances as a maximum conductance, e.g. \bar{g}_{Na} and \bar{g}_{K} , multiplied by a variable that represents the fraction of the maximum conductance actually open. The value of this variable is always between zero and one, by definition, and it can be a function of one or more *gating* variables. Microscopically, these gating variables describe the state of *activating* and *inactivating* gating particles associated with each channel. Each gating particle can be either open or closed, depending on time and on the membrane potential. For a given channel to be open, all gating particles must be in their open state.

We consider first the potassium conductance. The potassium current is modeled as

$$I_{\text{K}} = \bar{g}_{\text{K}} n^4 (V - E_{\text{K}}) \quad (1.7)$$

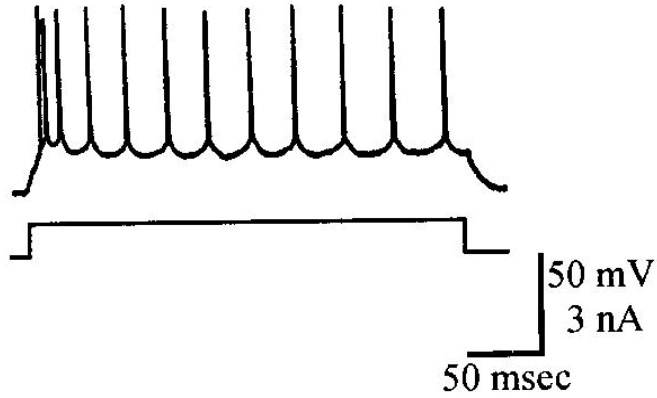


Figure 1.4: Response of a regular spiking neuron in cortex to a constant applied current (taken from [4]).

where n represents the state of the gating variable. Four gating particles are associated with each channel, hence the power of four. The dynamics of the gating variable is modeled by first order kinetics and the reaction scheme is



α_n is the rate of transitions from the closed to the open state, and β_n is the transition rate from the open to the closed state. They are voltage-dependent and have the dimensions of 1/sec. The corresponding first-order differential equation is

$$\frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V)n \quad (1.9)$$

Another way of expressing the last equation is by

$$\frac{dn}{dt} = \frac{n_\infty - n}{\tau_n} \quad (1.10)$$

where

$$n_\infty = \frac{\alpha_n}{\alpha_n + \beta_n} \quad (1.11)$$

and

$$\tau_n = \frac{\Phi}{\alpha_n + \beta_n} \quad (1.12)$$

The parameter Φ can be thought of as a scaling of the time units. We have taken $\alpha_n = -0.01(V + 34)/(\exp(-0.1(V + 34)) - 1)$ and $\beta_n = 0.125 \exp(-(V + 44)/80)$ and $\Phi = 10$.

The dynamics of the sodium conductance are more complex. They are described by an activation variable m and an *inactivation variable* h , and have the form

$$I_{\text{Na}} = \bar{g}_{\text{Na}} m^3 h (V - E_{\text{Na}}) \quad (1.13)$$

where \bar{g}_{Na} is the maximal sodium conductance and E_{Na} is the sodium reversal potential, which is typically positive, so that the sodium current is an inward current. The dynamical equations for m and h are similar those of n . We have taken $\alpha_m = -0.1(V + 30)/(\exp(-0.1(V + 30)) - 1)$, $\beta_m = 4 \exp(-(V + 55)/18)$, $\alpha_h = 0.07 \exp(-(V + 44)/20)$ and $\beta_h = 1/(\exp(-0.1(V + 14)) + 1)$.

The A-current

There is a large diversity of Potassium channels (for a review see [5]). We consider here the A-current (I_{A}), which we use in our single neuron conductance-based model. This current has also been called the fast transient K current, transient outward current (I_{to}), and rapidly inactivating K current.

The empirical model, first proposed by Neher [5], suggests the following form

$$I_{\text{A}} = \bar{g}_{\text{A}} a^3 b (V - E_{\text{K}}) \quad (1.14)$$

where a is a fast activation variable and b is a slow inactivation variable. Since the dynamics of a are very fast we replace it by a_{∞} where

$$a_{\infty} = 1/(\exp(-(V + 50)/20) + 1) \quad (1.15)$$

The dynamics of b are given by

$$db/dt = (b_{\infty} - b)/\tau_{\text{A}} \quad (1.16)$$

where

$$b_{\infty} = 1/(\exp((V + 80)/6) + 1) \quad (1.17)$$

The rates of opening and closing of K_{A} channels are reported to depend only weakly on the membrane potential. Thus, for the sake of simplicity we take the time constant, τ_{A} , to be voltage independent.

In the typical resting potential of the neuron most K_A channels are inactivated. The hyperpolarization after an action potential acts to remove the inactivation of the K_A channels. The depolarization that occurs afterwards increases a , and since b decays with a slow time constant the conductance of the A-current remains high for a while.

Synaptic transmission

Communication among neurons takes place at the *synapses*. In this work we consider only *Chemical synapses*, which are the main type of synapses in the adult mammalian central nervous system. When spiking activity occurs at the presynaptic cell, it causes the release of a *chemical transmitter*, which in turn binds to channels at the postsynaptic membrane and generates a transient increase of the conductance. The postsynaptic current of a single synapse has the form

$$I^{syn} = g_{syn}(t)(V(t) - E_{syn}) \quad (1.18)$$

where $g_{syn}(t)$ is the time course of the conductance change and E_{syn} is the reversal potential of the ion that flows through the postsynaptic channels, known as the *synaptic reversal potential*.

Typically, the conductance change that accompanies a synaptic event has a fast rise to some peak value and a slow exponential decay, of the form

$$\frac{d}{dt}g_{syn}(t) = -\frac{g_{syn}(t)}{\tau} \quad (1.19)$$

where τ is the *synaptic time constant*.

Network dynamics

Finally, we can write the dynamic equations for a general network of N coupled single-compartment neurons

$$\begin{aligned} C \frac{dV_i(t)}{dt} &= -g_L(V_i(t) - E_L) - I_i^{active}(t) \\ &+ I_i^{app}(t) + I_i^{syn}(t), \quad (i = 1, \dots, N) \end{aligned} \quad (1.20)$$

where $V_i(t)$ is the membrane potential of the i -th cell at time t and $I_i^{active}(t)$ denotes the total sum of the active currents. An externally injected current is denoted by I^{app} . The synaptic current,

I^{syn} , is given by

$$I_i^{syn}(t) = \sum_{j=1}^N g_{ij}(t)(E_j - V_i(t)) \quad (1.21)$$

where $g_{ij}(t)$ is the synaptic conductance triggered by the action potentials of the presynaptic j -th cell, and E_j is the reversal potential of the synapse.

1.2.2 Rate Models

Conductance-based models of the form discussed above, provide a description of the underlying microscopical processes responsible for spike generation. In these models, the dynamical variables that describe the neuronal activities are the membrane voltages. Another approach is to characterize the state of the network units by smooth rate variables [6, 7]. Many rate models have the following form,

$$\tau_i \frac{dm_i}{dt} = -m_i + S(h_i), \quad i = 1, \dots, N \quad (1.22)$$

where m_i is the activity level of the i -th unit, τ_i is its integration time, and the function $S(x)$ describes the nonlinear input-output activation function of the unit. The input, h_i , is given by

$$h_i = \sum_{j=1}^N J_{ij}m_j + I_i, \quad i = 1, \dots, N \quad (1.23)$$

where the weight, J_{ij} , denotes the strength of the connection between the presynaptic unit, j , and the postsynaptic unit, i , and I_i denotes the external input to the i -th unit. When the external inputs, I_i , are constant in time, the above equations may have a stable static solution, which is a fixed point attractor of the dynamics, determined by:

$$m_i = S\left(\sum_{j=1}^N J_{ij}m_j + I_i\right), \quad i = 1, \dots, N \quad (1.24)$$

The simplicity of rate models makes them very useful for studying the properties of large networks. However, the biophysical interpretation of the parameters that appear in these models, such as J_{ij} , is not clear. Chapter 2 describes the framework we have developed to bridge between conductance-based models and rate models. It presents a precise mapping between the rate model parameters and those of the conductance-based model.

1.3 Optimization Principles in Modeling of Neuronal Networks

One plausible assumption about the organization of the nervous system is that it develops in such a way as to optimize the transmission and representation of information. As we all know, we are not perfect and not optimal in general. However, this assumption seems to apply very well at least for the early stages of sensory processing.

Efficiency of information representation may mean several things. On the one hand, the system has to preserve the existing information and take care not to lose or distort it. On the other hand, it is also desirable to sort out irrelevant information and extract the important features. Another point to consider is that the representation of information is not independent of the question of what reads the information, and what is to be done with this information. Different types of representations may prove useful for different computational tasks.

In practice, the idea is to construct objective functions that quantify the quality of information representation by a network of neurons. The assumption is that the network parameters, specifically, single neuron parameters and parameters that characterize the synaptic interactions, evolve in a manner that optimizes these objective functions. The major theoretical frameworks that are used for constructing such objective functions are information theory and statistical estimation theory.

Ideas concerning the use of information theoretic tools for modeling perceptual processing can be dated back to the works of Attneave (1954) and Barlow (1959). For instance, in an early paper by Attneave (1954) he writes:

A major function of the perceptual machinery is to strip away some of the redundancy of stimulation, to describe or encode information in a form more economical than that in which it impinges on the receptors.

Below, we review concepts from information theory and their application for constructing objective functions and learning algorithms for network parameters.

1.4 Information Theory

A thorough review of information theory can be found in [8]. Here we briefly go over the main concepts of information theory following the review in [2] (chapter 10).

Basic Concepts of Information Theory

Information

Consider a discrete random variable X that can take N values

$$X = \{x_k \mid k = 1, \dots, N\} \quad (1.25)$$

with probabilities

$$p_k = P(X = x_k), \quad k = 1, \dots, N \quad (1.26)$$

that fulfill the requirement

$$0 \leq p_k \leq 1 \quad \text{and} \quad \sum_{k=1}^N p_k = 1 \quad (1.27)$$

Each realization of X may be regarded as a *message*.

The amount of information that is gained after observing the event $X = x_k$ is defined as

$$I(x_k) = \log\left(\frac{1}{p_k}\right) = -\log(p_k) \quad (1.28)$$

$I(x_k)$ reflects the “uncertainty” that was removed with the observation, or equivalently, the amount of “surprise” when seeing the result. When the natural logarithm is used the units are *nats*, and when the base 2 logarithm is used the units are *bits*. The definition of information in Eq. (1.28) exhibits the following properties:

1.

$$I(x_k) = 0 \quad \text{for} \quad p_k = 1 \quad (1.29)$$

If we are certain about the outcome of an event, there is no information gained by its occurrence.

2.

$$I(x_k) \geq 0 \quad \text{for} \quad 0 \leq p_k \leq 1 \quad (1.30)$$

The occurrence of an event never results in a loss of information.

3.

$$I(x_k) > I(x_i) \quad \text{for} \quad p_k < p_i \quad (1.31)$$

The less probable an event is, the more information we gain through its occurrence.

Entropy

The amount of information $I(x_k)$ is in itself a discrete random variable with probability p_k . The mean value of $I(x_k)$ is called the entropy of the random variable X and is given by

$$\begin{aligned} H(X) &= E[I(x_k)] \\ &= \sum_{k=1}^N p_k I(x_k) \\ &= - \sum_{k=1}^N p_k \log(p_k) \end{aligned} \quad (1.32)$$

The meaning of entropy is the *average amount of information conveyed per message*. $H(X)$ is bounded as follows

$$0 \leq H(X) \leq \log(N) \quad (1.33)$$

and

1. $H(X) = 0$ if and only if $p_k = 1$ for some k , and the remaining probabilities in the set are all zeros. This lower bound on entropy corresponds to *no uncertainty*.
2. $H(X) = \log(N)$ if and only if $p_k = 1/N$ for all k (uniform distribution). This upper bound on entropy corresponds to *maximum uncertainty*.

Relative Entropy

The *relative entropy* or *Kullback-Leibler divergence (distance)* of two probability functions $p(x)$ and $q(x)$ of the random variable X is defined by

$$D_{p||q} = \sum_x p(x) \log \left(\frac{p(x)}{q(x)} \right) \quad (1.34)$$

where the sum is over all possible values that X can take. The relative entropy expresses the statistical distance between two distributions. It is non-negative and becomes zero only for the case of $p = q$. Note however, that this is not a symmetric measure.

Mutual Entropy

Given a joint distribution of two random variables, the *mutual entropy* is defined as

$$H(X, Y) = - \sum_{x,y} p(x, y) \log(p(x, y)) \quad (1.35)$$

It expresses the amount of uncertainty before the observation of X and Y .

Conditional Entropy

The *conditional entropy* $H(Y | X)$ measures the average amount of uncertainty on Y that is left after a measurement of X . It is defined by

$$\begin{aligned} H(Y | X) &= \sum_x p(x) H(Y | X = x) \\ &= - \sum_x p(x) \sum_y p(y | x) \log p(y | x) \\ &= - \sum_{x,y} p(x, y) \log(p(x | y)) \end{aligned} \quad (1.36)$$

and has the property that

$$0 \leq H(Y | X) \leq H(Y) \quad (1.37)$$

The conditional entropy can be written as

$$H(Y | X) = H(X, Y) - H(X) \quad (1.38)$$

Or in words, the total uncertainty related to X and Y minus the uncertainty related to X .

Mutual Information

$H(Y)$ represents the uncertainty about Y *before* observing X , and $H(Y | X)$ represents the uncertainty about Y *after* observing X . Thus, the difference $H(Y) - H(Y | X)$ represents our uncertainty about Y which is resolved by observing X . This quantity is called the *mutual information* between the random variables X and Y , and is denoted by $I(X, Y)$. Using Eq. (1.38) we can express it in the following forms

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) \\ &= H(Y) - H(Y | X) \\ &= H(X) - H(X | Y) \end{aligned} \quad (1.39)$$

The mutual information is also the relative entropy between the joint distribution $p(x, y)$ and the product $p(x)p(y)$

$$I(X, Y) = \sum_{x,y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (1.40)$$

Therefore, it is always positive and equals zero if and only if X and Y are statistically independent.

Information Theoretic Concepts for Continuous Random Variables

The *differential entropy* of a continuous random variable X with *probability density function* $f(x)$, is defined by

$$\begin{aligned} h(X) &= - \int_{-\infty}^{\infty} f(x) \log f(x) dx \\ &= -E [\log f(x)] \end{aligned} \quad (1.41)$$

For justification see [2].

The mutual information between two continuous random variables X and Y is given by

$$\begin{aligned} I(X, Y) &= h(X) + h(Y) - h(X, Y) \\ &= h(Y) - h(Y | X) \\ &= h(X) - h(X | Y) \end{aligned} \quad (1.42)$$

where the *conditional differential entropy* of Y given X is defined by

$$h(Y | X) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \log f(y | x) dx dy \quad (1.43)$$

Maximum Mutual Information Principle

In 1988 Linsker [9] proposed the following principle (taken here from [2]):

The transformation of a random vector \mathbf{X} observed in the input layer of a neural system to a random vector \mathbf{Y} produced in the output layer of the system should be so chosen that the activities of the neurons in the output layer jointly maximize information about the activities in the input layer. The objective function to be maximized is the mutual information $I(\mathbf{X}, \mathbf{Y})$ between the vectors \mathbf{X} and \mathbf{Y} .

This principle is usually termed the *infomax* principle.

If the mapping from input to output is deterministic up to some additive noise, the conditional entropy $H(Y | X)$ is simply the entropy of the noise and does not depend on network parameters. In this situation, the basic quantity to be maximized is simply the entropy of the outputs.

1.5 Undercomplete, Complete and Overcomplete Representations

A multi-dimensional input can be represented by a smaller, equal or larger number of components. Correspondingly, the representation is termed *undercomplete*, *complete* or *overcomplete*.

Efficient representation schemes seek to provide a good approximation of the statistical density of the data set. In undercomplete representations, the input is directly compressed, and the target is to find an optimal compression scheme. This approach is suitable when the inputs are generated as different combinations of a *small* number of basic patterns, or features. Learning algorithms for undercomplete representations try to extract these principal features, and represent each input by the corresponding coefficients. A well-known method for this is the PCA (Principal Component Analysis) [10].

In many natural data sets the variety of basic features is very large, while the number of components of each single data sample is not necessarily large. Overcomplete representations can capture this data structure by letting each output component represent a different feature. Typically, a small number of basic features will be present in a single data point. Thus, in response to a single input, a relatively small number of outputs will be active, resulting in a *sparse representation* of the input. Specifying only those outputs that are active give an efficient representation of the data. In general, overcomplete representations are *redundant*; namely the decomposition of a signal is not unique. However, an advantage of that, is that the representation is more robust, and can be more stable to small perturbations in the input.

1.6 Independent Component Analysis

Learning algorithms that can learn data representations from samples of the data set, have been a subject of great interest in recent years. One popular method for this task is Independent

Component Analysis (ICA). The standard ICA formulation assumes that the signals are generated by a linear mixing of *independent sources*. It also assumes that the number of signal components equals the number of sources; namely, that the representation is *complete*.

For instance, assume that two people are speaking in a room and two microphones, held in different locations, record their conversation. Each of the recorded time signals, $x_1(t)$ and $x_2(t)$, is a weighted sum of the speech signals emitted by the two speakers, $s_1(t)$ and $s_2(t)$,

$$x_1(t) = a_{11}s_1 + a_{12}s_2 \quad x_2(t) = a_{21}s_1 + a_{22}s_2 \quad (1.44)$$

The parameters a_{ij} depend on the distances of the microphones from the speakers. Estimating the two original speech signals using only the two recorded signals is called the *cocktail-party problem* or, more generally, *blind source separation*. To solve this, the ICA approach assumes that the two sources, $s_1(t)$ and $s_2(t)$, are *statistically independent*. This is a simplifying assumption, and is not true in general, since the speakers interact with each other.

The ICA problem can be solved on the basis of minimizing or maximizing several cost functions. One formulation of the problem uses the *maximum mutual information principle*, discussed in 1.4. The input to the network is a vector \mathbf{X} of N components. The outputs are of the form $g_i(\mathbf{W}_i^T \mathbf{X})$, ($i = 1, \dots, N$), where g_i are some non-linear scalar functions, and the \mathbf{W}_i are the weight vectors of the neurons. The target is to maximize the mutual information between input and output with respect to the \mathbf{W}_i vectors. Since the mapping from input to output is deterministic, the problem reduces to maximizing the entropy of the outputs (see section 1.4).

In Chapter 3 we introduce a generalization of this approach to *overcomplete* representations. We also allow recurrent interactions between the neurons at the output layer. The ICA learning algorithm is recovered from our algorithm as a special case, when the number of outputs and inputs is equal and the recurrent interactions are taken to be zero. A comprehensive review on the subject of ICA is given in [11].

1.7 Orientation Selectivity in Primary Visual Cortex

1.7.1 The visual pathway and receptive fields

Visual processing begins in the retina. The retinal ganglion cells (RGCs), which constitute the output layer of the retina, send their axons via the optic nerve fibers to the left and right Lateral Geniculate Nuclei (LGNs) in the thalamus. The thalamic afferents from the LGN terminate in layers 4 and 6 of the primary visual cortex (see Fig. 1.1).

Receptive fields Every cell along the above visual pathway has a *receptive field* (RF): an area in the visual field (or equivalently, on the retinal surface), such that presenting a stimulus in it generates a change in the firing rate of the neuron. In general, a uniform stimulus covering the entire RF will not generate any change in the response of the cell, and the cell will continue firing at its spontaneous rate. Only a luminance *gradient* that falls within the RF will cause the firing rate to change. RFs in the early stages of visual processing can generally be divided into two subfields: an *excitatory subfield*, which if stimulated with a small spot of light, that is brighter than its surroundings, generates an increase in the firing rate of the cell, and an *inhibitory subfield*, where the same spot of light generates a decrease in the firing rate. A dark spot will generate a decrease in the firing rate when stimulating the excitatory subfield, and an increase when stimulating the inhibitory one.

Retinal Ganglion Cells RGCs have concentric center-surround receptive fields. ON-center RGCs are those in which the excitatory subfield constitutes the center of their RF, and the inhibitory subfield lies in the ring around it. In OFF-center RGCs the opposite is true. Thus, RGCs respond optimally to round stimuli that fill the central subfield. The RFs of RGCs in or near the fovea, which corresponds to the center of gaze, are much smaller than those of cells many degrees out from the fovea. The smallest RF centers subtend about 2 minutes of the arc, the biggest can have diameters of 1 degree or more. The full RFs are several times bigger. By comparison, the moon, seen from earth, subtends one-half a degree.

LGN cells The RF properties of LGN cells do not differ substantially from those of RGCs. Geniculate neurons also have concentrically arranged RFs, of either the ON-center or OFF-center type, but have more equal matching of the excitatory and inhibitory subfield areas, resulting in weaker

responses to diffuse illumination.

1.7.2 Orientation selective cells

Cortical neurons exhibit responses to visual stimuli which are dramatically different from those exhibited earlier along the visual pathway. As discovered by Hubel and Wiesel, these cells no longer respond to small spots of light, but instead prefer elongated edges or bars with a specific orientation [12]. This specificity means that a neuron continues firing at its spontaneous rate of several spikes per second in response to a vertical bar, for instance, but elevates its firing rate up to 50 fold in response to a horizontal bar (see an example in Fig. 1.5). The stimulus orientation to which the neuron responds best is call the *preferred orientation* (PO) of the neuron.

Tuning curves Orientation selectivity is quantitatively measured via the cell's *tuning curve*, which is a plot of the average firing rate of the cell as a function of the stimulus orientation. The tuning curve peaks by definition at the PO, and the narrower it is, the stronger the selectivity of the neuron. A typical *tuning width*, defined as half width at half height, is 20° . The height of the tuning curve depends on the contrast of the stimulus - the higher the contrast the larger the response of the cell, but the *width* of the tuning curve is found to be independent of the contrast, as evident in Fig. 1.5 [13, 14].

Spatial organization and hypercolumns

Retinotopic organization The topographical arrangement of the cells in the retina is replicated in the LGN and in V1, and the visual field is thus mapped onto the cortical surface. This means that anatomically neighboring neurons have largely overlapping RFs. The amount of overlap between the RFs of two neurons is negatively correlated with the distance between the cortical columns they belong to. When the distance is larger than about 2 mm, usually there is no overlap between the RFs.

Orientation columns Another feature is the tendency of neighboring neurons in V1 to have similar POs. Neurons which share the same cortical column have roughly the same PO (see Fig. (1.6)). Moreover, when measuring the POs of cells situated along a line parallel to the cortical surface, one finds that they tend to change monotonically. The shift is about 10° for each $50 \mu\text{m}$, which

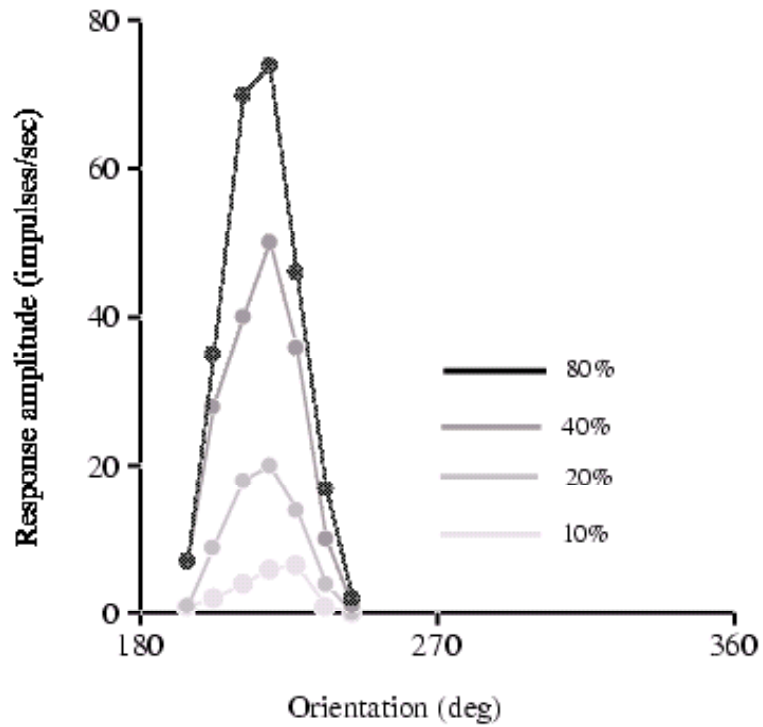


Figure 1.5: Demonstration of the contrast invariance effect. Shown are tuning curves of a V1 cell, with a preferred orientation of approximately 220° , in response to stimuli with different contrast values (taken from [13]).

means that we expect to find all possible orientations in a cortical tissue 1 mm wide. However, the PO change is occasionally non-monotonic and can even be non-continuous, with shifts of 45 to 90 degrees.

Ocular dominance columns The columns for eye preference constitute a second system of cortical organization. Because geniculate cells are monocular, any individual geniculate axon obviously belongs to the left eye or the right eye. It turns out that every geniculate fiber entering V1 terminates in layer 4C in clusters of synaptic endings, 0.5 mm wide, separated by blank areas of the same width. Thus, in the layer where the geniculate afferents terminate, all simple cells are stimulated by one eye only. Outside this layer, the simple and complex cells are driven by both eyes, but most of them are dominated to some degree by one eye. The ocular dominance columns form slabs 0.5 mm wide.

Hypercolumns The functional unit of V1 appears to be a roughly cubical aggregate of cells receiving input from both eyes, in which all the possible orientations are represented and all cells have overlapping RFs. Hubel and Wiesel first named this unit an *ice cube*, but it is usually named a *hypercolumn*. Although much oversimplified, the concept of the hypercolumn proves useful when studying the cortical microcircuitry. Its dimensions are on the order of one square millimeter, which corresponds to twice the width of the ocular dominance columns in one dimension, and allows for a full representation of all POs along the other dimension (see Fig. (1.6)). Note that the dimensions of the hypercolumn correspond to the dimensions of the local network described earlier.

1.7.3 Theories of orientation selectivity

A comprehensive and concise review by Sompolinsky and Shapley of mechanisms for orientation selectivity appears in [1]. Here we briefly discuss the main points that are relevant for Chapter 4.

Hubel and Wiesel, who first described this phenomenon, proposed a pure feedforward mechanism according to which orientation selectivity is generated by the specific pattern of connections from the LGN to V1 [16]. Their model is supported by a large body of evidence, including recent experimental studies [17, 18]. However, the anatomical abundance of interactions among cortical neurons [20, 21, 22] and indirect physiological evidence for their contribution in the processing of visual stimuli [19], suggest that recurrent interactions in the cortex may play an important role in

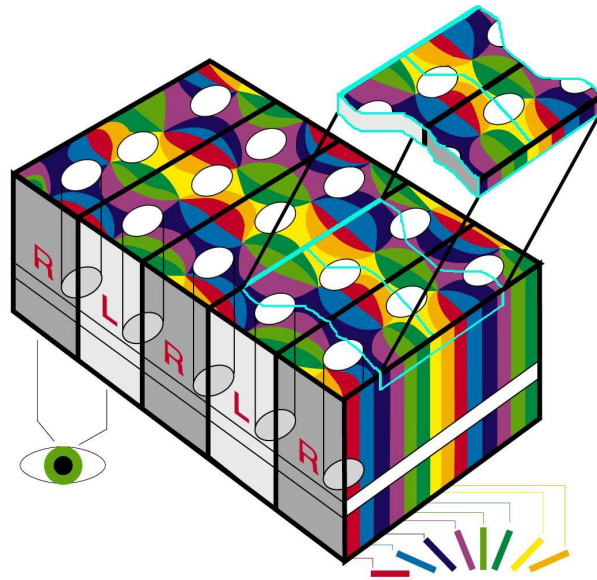


Figure 1.6: A schematic map of the functional organization of V1. Black lines mark the borders between columns of neurons that receive signals from the left and right eye and are responsible for depth perception. White ovals represent groups of neurons responsible for color perception (blobs). The “pinwheels” are formed by neurons involved in the perception of shape, with each color marking neurons responsible for a particular orientation in space. The top “slice” above the ice-cube model depicts two fundamental modules, each containing a complete set of about 60,000 neurons processing all the 3 features of orientation, depth and color (400 micron \times 800 micron). (Taken from the lab of Prof. Amiram Grinvald, The Weizmann Institute of Science).

this process. In addition, in the pure feedforward model, as the contrast of the stimulus increases, the width of the tuning increases as well. This cannot be reconciled with the experimental finding that the tuning width is independent of stimulus contrast [13].

Based on these findings, theoretical studies [23, 24] proposed that the selectivity to orientation involves recurrent cortical processing and emerges as a cooperative property of local networks in cortex. According to these models, the 'net' interactions between orientation columns have a "Mexican hat" shape: orientation columns that represent close orientations excite each other, while columns with a large angular separation inhibit each other. This pattern of interactions amplifies the input from the LGN at the stimulus orientation and suppresses it at nonoptimal orientations, thus increasing the sharpness of the tuning. When the modulation amplitude of the interactions is strong enough, the steady-states of the network dynamics form a continuum of attractors, differing one from each other by the location of the peak of the activity profile. Under these conditions, an activity profile peaked at some orientation will form even if the input from the LGN is not tuned. This kind of behavior is known as *spontaneous symmetry breaking*.

The neuronal tuning curves in these recurrent network models present contrast invariance. A high contrast results in increased positive feedback at the stimulus orientation, but also in increased negative feedback at nonoptimal orientations, thus keeping the width almost independent of the contrast.

The main question that arises is whether the mechanism underlying orientation selectivity is a feedforward mechanism or rather a recurrent one. To quote Sompolinsky and Shapley:

"Determining which of these two explanations is correct will have a big impact on our conception of cerebral cortical function. If orientation selectivity results from the specific pattern of convergence of LGN afferents, as proposed by Hubel and Wiesel, this would mean that a fundamental function of visual cortex arises from the feedforward filtering of sensory input, supporting the view of cortical processing as a hierarchy of feedforward transformations of neural representations. However, if intracortical circuitry plays a crucial role in orientation tuning, then this implies that intracortical dynamics shapes the internal representations of the external world in cortex."

Bibliography

- [1] Peretz L (1989). *Brain, structure and function: historical concepts*. Ministry of Defence Press, Tel Aviv.
- [2] Haykin S (1999). *Neural networks: a comprehensive foundation*. 2nd ed., Prentice-Hall, Upper Saddle River, NJ.
- [3] Tuckwell HC (1988). *Introduction to theoretical neurobiology: volume 2, nonlinear and stochastic theories*. Cambridge University Press, Cambridge.
- [4] Koch C (1999). *Biophysics of Computation: Information processing in Single Neurons*. Oxford University Press, New York.
- [5] Hille B (1984). *Ionic channels of excitable membranes*. Sinauer, Sunderland, MA.
- [6] Wilson HR & Cowan J (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J*, **12**:1–24.
- [7] Hopfield JJ (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci USA*, **79**:1554–2558.
- [8] Cover MT & Thomas JA (1991). *Elements of Information Theory*. 2nd ed., John Wiley and Sons.
- [9] Linsker R (1988). Self-organization in a perceptual network. *IEEE Computer*, **21**:105–117.
- [10] Jolliffe IT (1986). *Principal Component Analysis*. New York: Springer-Verlag.
- [11] Hyvärinen A, Karhunen J & Oja E (2001). *Independent Component Analysis*. John Wiley and Sons.
- [12] Hubel DH (1988) *Eye, brain, and vision*, Scientific American Library.
- [13] Sclar G & Freeman R (1982) Orientation selectivity in cat's striate cortex is invariant with stimulus contrast. *Exp Brain Res*, **46**:457–461.

- [14] Skottun B, Bradley A, Sclar G, Ohzawa I & Freeman R (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J Neurophysiol*, **57**:773–786.
- [15] Sompolinsky H & Shapley R (1997). New perspectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology*, **7**:514–522.
- [16] Hubel DH & Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture of the cat’s visual cortex. *J Physiol (Lond)*, **160**:106–154.
- [17] Reid RC & Alonso JM (1995) Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, **378**:281–284.
- [18] Ferster D, Sooyoung C & Wheat H (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, **380**:249–252.
- [19] Borg-Graham LJ, Monier C & Fregnac Y (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, **393**:369–373.
- [20] Martin KAC (1988). The Wellcome Prize Lecture: From single cells to simple circuits in the cerebral cortex. *Q J Exp Physiol*, **73**:637–702.
- [21] Douglas RJ & Martin KAC (1991) A functional microcircuit for cat visual cortex. *J Physiol*, **440**:735–769
- [22] Tarczy-hornoch K, Martin KAC, Stratford KJ & Jack JJB (1999). Intracortical excitation of spiny neurons in layer 4 of cat striate cortex in vitro. *Cereb Cortex*, **9**:833–843
- [23] Ben-Yishai R, Lev Bar-Or R & Sompolinsky H (1995). Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci USA*, **92**:3844–3848.
- [24] Somers DC, Nelson SB & Sur M (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neurosci*, **15**:5448–5465.

Chapter 2

Rate Models for Conductance Based Cortical Neuronal Networks

2.1 Abstract

Population rate models provide powerful tools for investigating the principles that underlie the cooperative function of large neuronal systems. However, biophysical interpretations of these models have been ambiguous. Hence, their applicability to real neuronal systems and their experimental validation have been severely limited. In this work, we show that conductance-based models of large cortical neuronal networks can be described by simplified rate models, provided that the network state does not possess a high degree of synchrony. We first derive a precise mapping between the parameters of the rate equations and those of the conductance-based network models for time independent inputs. This mapping is based on the assumption that the effect of increasing the cell's input conductance on its f-I curve is mainly subtractive. This assumption is confirmed by a single compartment Hodgkin-Huxley type model with a transient potassium A-current. This approach is applied to the study of a network model of a hypercolumn in primary visual cortex. We also explore extensions of the rate model to the dynamic domain, by studying the firing rate response of our conductance-based neuron to time dependent noisy inputs. We show that the dynamics of this response can be approximated by a time dependent second-order differential equation. This phenomenological single-cell rate model is used to calculate the response of a conductance-based

network to time dependent inputs.¹

2.2 Introduction

Theoretical models of the collective behavior of large neuronal systems can be divided into two categories. One category attempts to incorporate the known microscopic anatomy and physiology of the system. To study these models, numerical simulations are required. They involve a large number of parameters whose precise values are unknown, and the systematic exploration of the model parameter space is impractical. Furthermore, due to their complexity, it is hard to construct a qualitative interpretation of their behavior. The second category consists of simplified models that retain only some gross features of the modeled system, thereby allowing for systematic analytical and numerical investigations. These models have been extremely useful in extracting qualitative principles underlying such functions as memory, visual processing and motor control (Amit, 1989; Churchland & Sejnowski, 1992; Georgopoulos & Lukashin, 1993; Ben-Yishai, Lev Bar-Or & Sompolinsky, 1995; Seung, 1996; Zhang, 1996; Salinas & Abbott, 1996; Hansel & Sompolinsky, 1998; Rolls & Treves, 1998).

Simplified models of large neuronal systems are often cast in the form of rate models, in which the state of the network units is characterized by smooth rate variables (Wilson & Cowan, 1972; Hopfield, 1984). These variables are related to the units' synaptic inputs via a nonlinear input-output transfer function. The input is a linear sum of the presynaptic activities, whose coefficients are termed the "synaptic weights" of the network. Unfortunately, the use of rate models for concrete neuronal systems has been limited by the lack of a clear biophysical interpretation of the parameters appearing in these models. In particular, the relation between activity variables, input variables, and synaptic weights, on one hand, and physiologically measured quantities on the other, is obscure. Furthermore, quite often rate models predict that the network should settle in a fixed point where the network activities as well as synaptic inputs are time independent. However, the biological meaning of this fixed point state is unclear, since neither the postsynaptic currents nor the postsynaptic potentials are constant in time if the cells are active. It is thus important

¹This chapter appeared as: Shriki O, Hansel D, and Sompolinsky H (2003). Rate Models for Conductance Based Cortical Neuronal Networks. *Neural Computation*, **15(8)**:1809-1841.

to inquire whether there is a systematic relation between real neuronal systems and simple rate models.

Several studies have derived reduced rate models for networks of spiking neurons (Amit & Tsodyks, 1991; Abbott and Kepler, 1990; Ermentrout, 1994). In particular, it has been shown that if the synaptic time constants are long the network dynamics can be reduced to rate equations which describe the slow dynamics of the synaptic activities (Ermentrout, 1994). However, the assumption of slow synaptic dynamics is inadequate for modeling cortical networks, where fast synapses play a dominant role. Here, we show that asynchronous states of large cortical networks described by conductance-based dynamics can be described in terms of simple rate equations, even when the synaptic time constants are small. A simple mapping between the synaptic conductances and the synaptic weights is derived. We apply our method to study the properties of conductance based networks that model a hypercolumn in visual cortex. The simple reduction of conductance based networks to rate models is restricted to asynchronous states which exist only if the networks are driven by stationary inputs. We derive a more complex rate model which is appropriate to describe the synchronous response of large conductance based networks to weakly non-stationary noisy synaptic inputs. Our results provide a framework for using rate models to quantitatively predict the extracellular and intracellular response properties of large cortical networks.

2.3 Models and Methods

2.3.1 Dynamics of a single-compartment cell

Our starting point is the dynamic equation of a single-compartment neuron,

$$C \frac{dV(t)}{dt} = g_L(E_L - V(t)) - I^{active}(t) + I^{app}(t) \quad (2.1)$$

where $V(t)$ is the membrane potential of the cell at time t , C is its capacitance, g_L is the leak conductance and E_L is the reversal potential of the leak current. Besides the leak current, the cell has active ionic currents with Hodgkin-Huxley type kinetics (Hodgkin and Huxley, 1952), the total sum of which is denoted as $I^{active}(t)$ in Eq. 2.1. An externally injected current is denoted as I^{app} . If I^{app} is constant in time and is sufficiently large, the cell will fire in a repetitive manner with a steady state firing rate f . In general, the relation between the applied current, I , and the

firing rate, f , defines a function $f = F(I, g_L)$, called the *f-I curve*. The second argument, g_L , expresses the dependence of the input-output function of the neuron on the magnitude of the leak conductance. This dependence is an important factor in our work as will become clear later. The form of the function F depends on the active currents comprising I^{active} . In many cortical neurons the f-I curve is approximately linear for I above threshold (Azouz, Gray, Nowak, & McCormick, 1997; Ahmed, Anderson, Douglas, Martin, & Whitteridge, 1998; Stafstrom, Schwindt, & Crill, 1984) and can be captured by the following equation

$$f = \beta[I - I_c]_+ \quad (2.2)$$

where $[x]_+ \equiv x$ for $x > 0$ and is zero otherwise; β is the gain parameter. This behavior can be modeled by a Hodgkin-Huxley type single compartment neuron with a slow A-current (Hille, 1984). See Appendix A for the details of the model. The parameters of the sodium and the potassium currents were chosen to yield a saddle-node bifurcation at the onset of firing. Figure 2.1 shows the response of this model neuron without and with the A-current. As can be seen, the A-current linearizes the f-I relationship. Our model neuron has gain value $\beta = 35.4 \text{ cm}^2/\mu\text{Asec}$.

Relatively few experimental data have been published on the dependence of the firing rate of cortical cells on their leak conductance. However, experimental evidence (Conors, Malenka, & Silva, 1998; Brizzi, Hansel, Meunier, van Vreeswijk, & Zytnicki, 2001) and biophysical models (Kernell, 1968; Holt & Koch, 1997) show that increasing g_L affects the f-I curve primarily by increasing its threshold current, whereas its effect on the gain of the curve is weak. We incorporate these properties by assuming that β is independent of g_L and that the threshold current increases linearly with the leak conductance,

$$I_c = I_c^o + V_c g_L \quad (2.3)$$

The *threshold gain potential* V_c measures the rate of increase of the threshold current as the leak conductance, g_L , increases and I_c^o is the threshold current when $g_L = 0$. This behavior is also reproduced in our model neuron, as shown in Figure 2.2. Approximating the f-I curve with Eqs. 2.2-2.3 yields $I_c^o = 0.63 \mu\text{A}/\text{cm}^2$ and $V_c = 5.5 \text{ mV}$. This provides a good approximation of the f-I curve for the range $f = 5 - 150$ spikes/sec. For higher firing rates, the effect of the saturation of the rates becomes significant, and needs to be incorporated into the model. We have found that

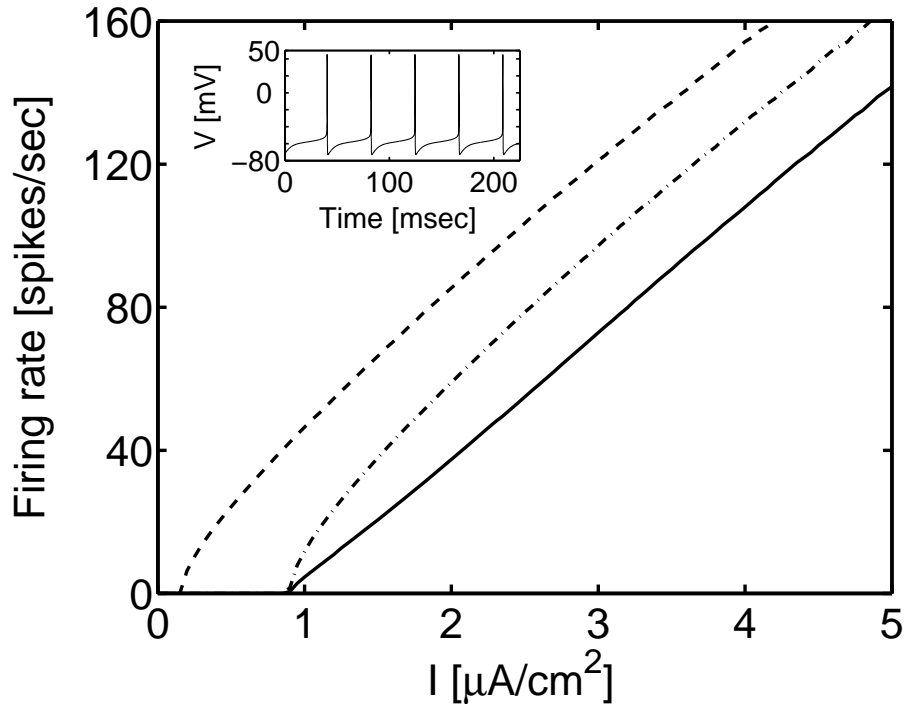


Figure 2.1: f-I curves of the single neuron model with $g_A = 0$ (dashed line), $g_A = 20 \text{ mS/cm}^2$, $\tau_A = 0 \text{ msec}$ (dash dotted line), $g_A = 20 \text{ mS/cm}^2$, $\tau_A = 20 \text{ msec}$ (solid line). Comparison of the three curves shows that linearization of the f-I curve is due to the long time constant of the A-current. Inset: A voltage trace of the single neuron model with constant current injection of amplitude $I = 1.6 \text{ } \mu\text{A/cm}^2$ for $g_A = 20 \text{ mS/cm}^2$, $\tau_A = 20 \text{ msec}$. The neuron's parameter values are as defined in Appendix A.

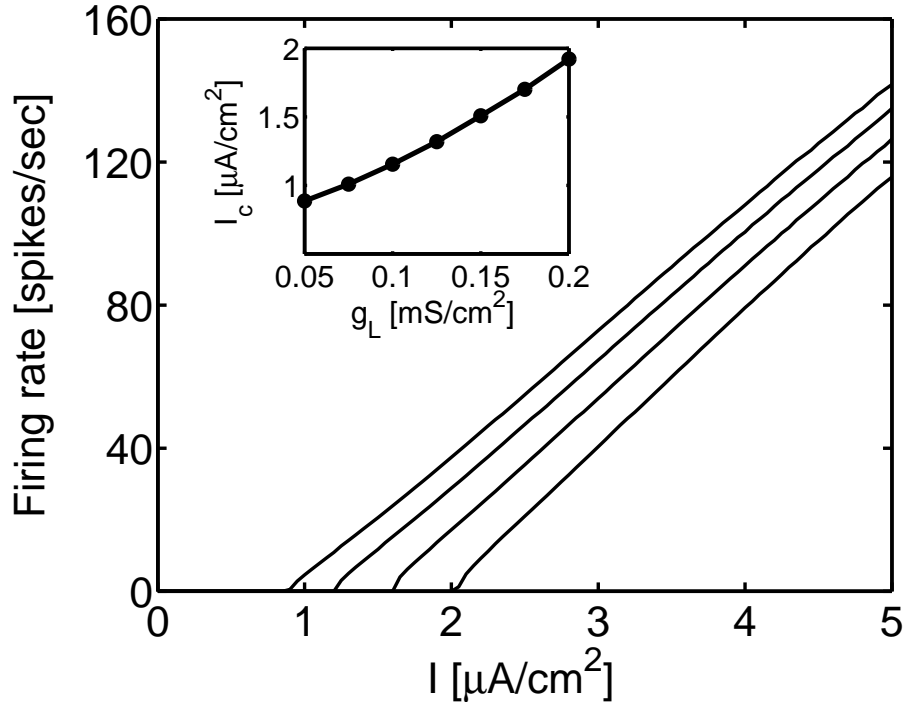


Figure 2.2: f-I curves for $g_A = 20 \text{ mS/cm}^2$, $\tau_A = 20 \text{ msec}$, and different values of g_L . The curves from left to right correspond to $g_L = 0.05, 0.1, 0.15, 0.2 \text{ mS/cm}^2$, respectively. Inset: the threshold current, I_c , as a function of the leak conductance, g_L .

this effect can be described approximately by an f-I curve of the form

$$f = \beta[I - I_c]_+ - \gamma[I - I_c]_+^2 \quad (2.4)$$

with I_c given by Eq. 2.3. Fitting the firing rate of our single neuron model to Eq. 2.4 yields good fit over the range $f = 5 - 300 \text{ spikes/sec}$, with the parameter values: $\beta = 39.6 \text{ cm}^2/\mu\text{Asec}$, $\gamma = 0.86 \text{ (cm}^2/\mu\text{A)}^2/\text{sec}$, $\alpha = 6.77 \text{ mV}$, and $I_c^o = 0.59 \mu\text{A/cm}^2$. The above conductance-based model neuron is the one used in all subsequent numerical analyses.

2.3.2 Network dynamics

The network dynamics of N coupled cells are given by

$$C \frac{dV_i}{dt} = g_L(E_L - V_i(t)) - I_i^{active} + I_i^{ext} + I_i^{net} \quad (i = 1, \dots, N) \quad (2.5)$$

where I_i^{net} denotes the synaptic current of the postsynaptic cell i generated by the presynaptic sources within the network. It is modeled as

$$I_i^{net}(t) = \sum_{j=1}^N g_{ij}(t)(E_j - V_i(t)) \quad (2.6)$$

where $g_{ij}(t)$ is the synaptic conductance triggered by the action potentials of the presynaptic j -th cell, and E_j is the reversal potential of the synapse. The synaptic conductance is assumed to consist of a linear sum of contributions from each of the presynaptic action potentials. In our simulations $g_{ij}(t)$ has the form of an instantaneous jump from 0 to G_{ij} followed by an exponential decay with a synaptic decay time τ_{ij} ,

$$\frac{dg_{ij}}{dt} = -\frac{g_{ij}}{\tau_{ij}} + G_{ij}R_j(t), \quad t > 0 \quad (2.7)$$

where $R_j(t) = \sum_{t_j} \delta(t - t_j)$ is the instantaneous firing rate of the presynaptic j -th neuron and t_j are the times of occurrence of its spikes. In general, we define G_{ij} as the peak of $g_{ij}(t)$ and the synaptic time constant as $\tau_{ij} = \int_0^\infty g_{ij}(t)dt/G_{ij}$.

Synaptic currents from presynaptic sources outside the network are denoted by I_i^{ext} . For simplicity, we assume that these sources are all excitatory with the same reversal potential, E^{inp} , peak conductance G^{inp} , and synaptic time constant, τ^{inp} . We assume that these sources fire Poisson trains of action potentials asynchronously, which generate synaptic conductances with dynamics similar to Eq. 2.7. Under these assumptions, their summed effects on the postsynaptic neuron i can be represented by a single effective excitatory synapse with peak conductance, G^{inp} , time constant τ^{inp} and activated by a single Poisson train of spikes with average rate f_i^{inp} which is the summed rate of all the external sources to the i -th neuron. Thus, the external current on neuron i can be written as:

$$I_i^{ext}(t) = g_i^{inp}(t)(E^{inp} - V_i(t)). \quad (2.8)$$

The quantity $g_i^{inp}(t)$ satisfies an equation similar to Eq. 2.7,

$$\frac{dg_i^{inp}}{dt} = -\frac{g_i^{inp}}{\tau^{inp}} + G^{inp}R_i^{inp}(t), \quad t > 0 \quad (2.9)$$

where R_i^{inp} is a Poisson spike-train with mean rate f_i^{inp} . The value of f_i^{inp} is specified below for each of the concrete models that we study. Due to the Poisson statistics of R_i^{inp} , the external conductance, $g_i^{inp}(t)$, is a random variable with a time average and variance $G^{inp}f_i^{inp}\tau^{inp}$ and

$(G^{\text{inp}})^2 f_i^{\text{inp}} \tau^{\text{inp}}/2$, respectively. Note that $f_i^{\text{inp}} \tau^{\text{inp}}$ is the mean number of input spikes arriving within a single synaptic integration time. The fluctuations in R_i^{inp} constitute the noise in the external input to the network. We define the coefficient of variation of this noise as the ratio of its standard deviation and its mean, i.e. $\Delta_i = 1/\sqrt{2f_i^{\text{inp}} \tau^{\text{inp}}}$. As expected the coefficient of variation is proportional to the inverse square root of the total number of spikes arriving within a single synaptic integration time. In particular, one can increase the noise level of the input by decreasing f_i^{inp} and increasing G^{inp} while keeping their product constant.

Time dependent inputs are modeled by a Poisson process with a rate which is modulated in time. In most of the examples studied in this paper, we assume a sinusoidal modulation, namely that the instantaneous firing rate in the external input to the neuron is:

$$f^{\text{inp}}(t) = f_0^{\text{inp}} + f_1^{\text{inp}} \cos(\omega t) \quad (2.10)$$

where $\omega/2\pi$ is the frequency of the modulation.

2.3.3 Model of a hypercolumn in primary visual cortex

We model a hypercolumn in visual cortex by a network consisting of N_e excitatory neurons and N_{in} inhibitory neurons that are selective to the orientation of the visual stimulus in their common receptive field. We impose a *ring* architecture on the network. The cortical neurons are parameterized by an angle θ , which denotes their preferred orientation (PO). The i -th excitatory neuron is parameterized by $\theta_i = -\frac{\pi}{2} + i\frac{\pi}{N_e}$, and similarly for the inhibitory ones. The peak conductances of the cortical recurrent excitatory and inhibitory synapses decay exponentially with the distance between the interacting neurons, measured by the dissimilarity in their preferred orientations, i.e.,

$$G_\alpha(\theta - \theta') = \frac{\bar{G}_\alpha}{\lambda_\alpha} \exp(-|\theta - \theta'|/\lambda_\alpha) \quad (2.11)$$

where $\theta - \theta'$ is the difference between the POs of the pre and postsynaptic neurons. The index α takes the values e and in . The quantity G_e (resp. $\alpha = in$) denotes an excitatory (inhibitory) interaction (targeting either excitatory or inhibitory neurons) with a space constant λ_e (resp. λ_{in}). Note that the excitatory as well as the inhibitory interactions are the same for excitatory and inhibitory targets. Additional excitatory neurons provide external input to the network, representing the

LGN input to cortex, each with peak conductance $G^{inp} = \bar{G}_{LGN}$. The total mean firing rate of the afferent inputs to a neuron with PO θ is $f^{inp} = f_{LGN}(\theta - \theta_0)$, where

$$f_{LGN}(\theta - \theta_0) = \bar{f}_{LGN}C[(1 - \epsilon) + \epsilon \cos(2(\theta - \theta_0))] \quad (2.12)$$

The parameter C is the stimulus contrast, and the angle θ_0 denotes the orientation of the stimulus. The parameter ϵ measures the degree of tuning of the LGN input. If $\epsilon = 0$, the LGN input is untuned: all the neurons receive the same input from the LGN, regardless of their PO and the orientation of the stimulus. If $\epsilon = 0.5$, the LGN input vanishes for neurons with a PO which is orthogonal to the stimulus orientation. The maximal LGN rate \bar{f}_{LGN} is the total firing rate of the afferents of a stimulus with $C = 1$ and $\theta = \theta_0$. The single neuron dynamics is given by Eq. 2.1, and is assumed to be the same for both the excitatory and inhibitory populations.

2.3.4 Numerical integration and analysis of spike responses

In the numerical simulations of the conductance-based networks the non-linear differential equations of the neuronal dynamics were integrated using a fourth-order Runge-Kutta method (Press, Flannery, Teukolsky & Vetterling, 1988) with a fixed time step Δt . Most of the simulations were performed with $\Delta t = 0.05$ msec. In order to check the stability and precision of the results some simulations were also performed with $\Delta t = 0.025$ msec.

A spike event is counted each time the voltage of a neuron crosses a fixed threshold value $V_{th} = 0$. We measure the instantaneous firing rate of a single neuron defined as the number of spikes in time bins of size $\Delta t = 0.05$ msec, averaged over different realizations of the external input noise. We then compute the time average of this response, and in the case of periodically modulated input, the amplitude and phase of its principal temporal harmonic. We also measure the population firing rate, defined as the number of spikes of single neurons in each time bin divided by Δt , averaged over a select population of neurons in the network as well as over the input noise. As in the case of a single neuron, the network response is characterized by the time average and the principal harmonic of the population rate.

2.4 Rate Equations for General Asynchronous Neuronal Networks

The dynamic states of a large network characterized by the above equations can be classified as being synchronous or asynchronous (Ginzburg & Sompolinsky, 1994; Hansel & Sompolinsky, 1996), which differ in terms of the strength of the correlation between the temporal firing of different neurons. When the external currents, I_i^{ext} , are constant in time (except for a possible noisy component which is spatially uncorrelated), the network may exhibit an asynchronous state in which the activities of the neurons are only weakly correlated. Formally, in an asynchronous state the correlation coefficients between the voltages of most of the neuronal pairs approach zero in the limit where the network size, N , grows to infinity.

Analyzing the asynchronous state in a highly connected network is relatively simple. Because each post synaptic cell is affected by many uncorrelated synaptic conductances (within a window of its integration time), these conductances can be taken to be time independent. In other words, in the asynchronous state the spatial summation of the synaptic conductances is equivalent to a time average. Hence, the total synaptic current of each cell, can be written as

$$I_i^{net}(t) + I_i^{ext}(t) = \sum_{j=1}^N G_{ij} \tau_{ij} f_j(E_j - V_i(t)) + G^{inp} \tau^{inp} f_i^{inp}(E^{inp} - V_i(t)) , \quad (2.13)$$

see Eqs. 2.6 and 2.8. This current can be decomposed into two components:

$$I_i^{net} + I_i^{ext} = I_i^{app} + \Delta I_i^L . \quad (2.14)$$

I_i^{app} is the component of the synaptic current that has the form of a constant applied voltage-independent current

$$I_i^{app} = \sum_{j=1}^N G_{ij} \tau_{ij} f_j(E_j - E_L) + G^{inp} \tau^{inp} f_i^{inp}(E^{inp} - E_L) \quad (2.15)$$

The second component of the synaptic current, ΔI_i^L , embodies the voltage dependence of the synaptic current and has the form of a leak current

$$\Delta I_i^L = g_i^{syn} (E_L - V_i(t)) , \quad (2.16)$$

where

$$g_i^{syn} = g_i^{net} + g_i^{inp} \quad (2.17)$$

is the mean total synaptic conductance of the i -th cell. The quantities g_i^{net} and g_i^{inp} are given by

$$g_i^{net} = \sum_{j=1}^N G_{ij} \tau_{ij} f_j , \quad (2.18)$$

$$g_i^{inp} = G^{inp} \tau^{inp} f_i^{inp} . \quad (2.19)$$

Thus, the discharge of the postsynaptic cell in the asynchronous network can be described by the f-I curve of a *single* cell with an applied current, Eq. 2.15, and a 'leak' conductance, which is equal to $g_L + g_i^{syn}$, Eq. 2.17. Incorporating these contributions in Eq. 2.2, taking into account the dependence of the threshold current on the total passive conductance as given by Eq. 2.3, yields the following equations for the firing rates of the cells

$$\begin{aligned} f_i &= \beta [I_i^{app} - V_c(g_L + g_i^{syn}) - I_c^o]_+ \\ &= \beta \left[\sum_{j=1}^N J_{ij} f_j + J^{inp} f_i^{inp} - T \right]_+ , \quad (i = 1, \dots, N), \end{aligned} \quad (2.20)$$

where

$$J_{ij} = G_{ij} \tau_{ij} (E_j - E_L - V_c) \quad (2.21)$$

and

$$J^{inp} = G^{inp} \tau^{inp} (E^{inp} - E_L - V_c) \quad (2.22)$$

The parameter T is the threshold current of the isolated cells, i.e., $T = I_c(g_L)$. Note that the subtractive term V_c in Eqs. 2.21-2.22 is the result of the increase of the current threshold of the cell due to the synaptic conductance, see Eq. 2.3.

Equation 2.20 is of the form of the self-consistent rate equations that describe the input-output relations for the neurons in a recurrent network at a fixed point state. This theory provides a precise mapping between the biophysical parameters of the neurons and synapses and the parameters appearing in the fixed point rate equations. The output state variables, given by the right hand side of Eq. 2.20 are simply the stationary firing rates of the neurons. The input variables, $\sum_{j=1}^N J_{ij} f_j + J^{inp} f_i^{inp}$, are the mean synaptic currents at a fixed potential given by $E_L + V_c$, where V_c is the threshold-gain potential, Eq. 2.3.

Equation 2.21 provides a precise interpretation of the synaptic efficacies J_{ij} in terms of the biophysical parameters of the cells and the synaptic conductances. We note in particular that our

theory yields a precise criterion for the *sign* of the efficacy of the synaptic connection. According to Eq. 2.21 synapses with positive efficacy obey the inequality

$$E_j > E_L + V_c \quad (2.23)$$

Conversely, synapses with negative efficacies obey $E_j < E_L + V_c$. The potential $E_L + V_c$ is close but not identical to the threshold potential of the cell. Hence, this criterion which takes into account the dynamics of firing rates in the network, does not match exactly the biophysical definition of excitatory and inhibitory synapses.

The above results allow the prediction not only of the stationary rates of the neurons but also their mean synaptic conductances due to the input from within the network and to the external input. In fact, using Eqs. 2.18-2.19 and Eqs. 2.21-2.22 yields

$$g_i^{net} = \sum_{j=1}^N J_{ij} \frac{f_j}{E_j - E_L - V_c} \quad (2.24)$$

and

$$g_i^{inp} = J_i^{inp} \frac{f_i^{inp}}{E_i^{inp} - E_L - V_c} \quad (2.25)$$

In the following sections we apply this theory to concrete network architectures.

2.5 Response of an Excitatory Population to a Time-independent Input

We first test the mapping equations 2.20-2.22 in the case of a large homogeneous network that contains N identical excitatory neurons. The network dynamics are given by Eqs. 2.5-2.7 with the single neuron model of Appendix A. Each neuron is connected to all other neurons with a peak synaptic conductance, G , which is the same for all the connections in the network. In addition, each neuron receives a single external synaptic input that has a peak conductance, G^{inp} , which is activated by a Poisson process with a fixed uniform rate, f^{inp} . The external synaptic inputs to different cells are uncorrelated. The dynamic response of all synaptic conductances is given by Eq. 2.7 with a single synaptic time constant $\tau_e = 5$ msec.

Applying Eqs. 2.20-2.22 to this simple architecture results in the following equation for the mean firing rate of the neurons in the network, f ,

$$f = \beta \left[J^{\text{inp}} f^{\text{inp}} + Jf - I_c \right]_+ \quad (2.26)$$

where

$$\begin{aligned} J^{\text{inp}} &= G^{\text{inp}} \tau_e (E_e - E_L - V_c) \\ J &= NG \tau_e (E_e - E_L - V_c) \end{aligned} \quad (2.27)$$

The solution for the firing rate is

$$f = \frac{\beta}{1 - \beta J} \left[J^{\text{inp}} f^{\text{inp}} - I_c \right]_+ \quad (2.28)$$

The mean firing rate, f , of the neurons in the network, as predicted from this equation, is displayed in Figure 2.3 (dashed line) against the value of the peak conductance of the total excitatory feedback, NG . When the synaptic conductance increases, such that J reaches the critical value $J = 1/\beta$ (corresponding to $NG = 0.095 \text{ mS/cm}^2$), the firing rate, f , diverges. However, it is expected that when the firing rate reaches high values, the weak nonlinearity of the f-I curve, given by the quadratic correction Eq. 2.4, will need to be taken into account. Indeed, solving the self-consistent equation for f with the quadratic term (solid line in Figure 2.3) yields finite values for f , even when J is larger than $1/\beta$. In addition, the quadratic nonlinearity predicts that in the high J regime, the network should develop bistability. For a range of subthreshold external inputs, the network can be either in a stable quiescent state or in a stable active state with high firing rates, as shown in the Inset of Figure 2.3. These predictions are in full quantitative agreement with the numerical simulations of the conductance based excitatory network as shown in Figure 2.3.

2.6 A Model of a Hypercolumn in Primary Visual Cortex

In this section we show how the correspondence between conductance-based models and rate models can be applied to investigate a model of a hypercolumn in V1.

When applying the general rate equations 2.20-2.22 to the hypercolumn model, we first note that in the asynchronous state the firing rate profile of the excitatory and the inhibitory populations

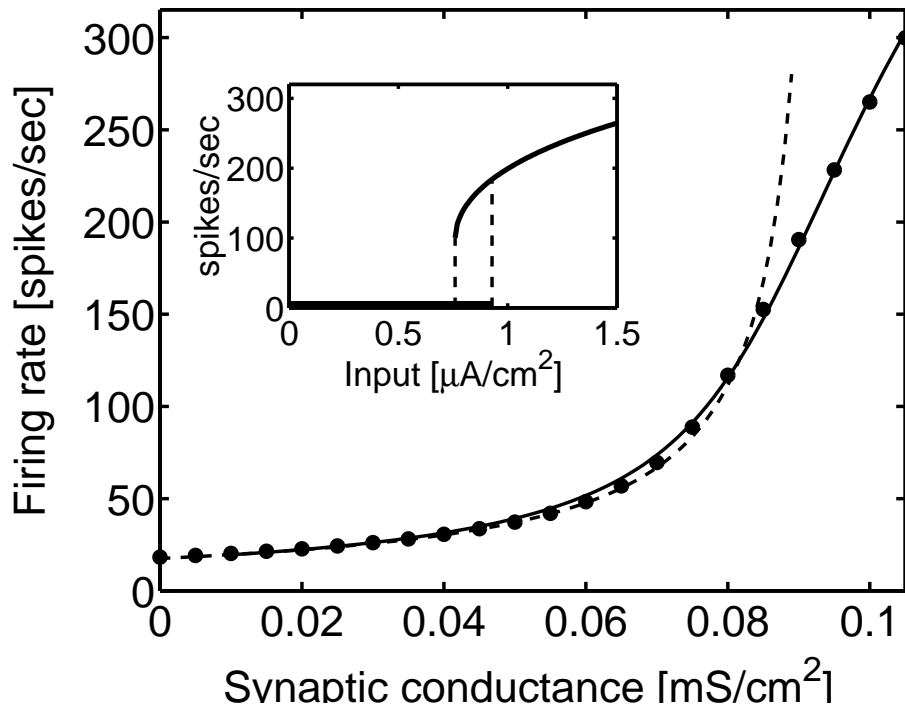


Figure 2.3: The firing rate versus excitatory synaptic strength in a large network of fully connected excitatory neurons. The rate of the external input is $f^{\text{inp}} = 1570$ spikes/sec and the synaptic time constant is $\tau_e = 5$ msec. Dashed line: analytical results from Eq. 2.28. Solid line: analytical results when a quadratic fit is used for the f-I curve, Eq. 2.4. Circles: results from simulations of the conductance-based model with $N = 1000$. Inset: The firing rate versus external input for strong excitatory feedback (analytical results) showing bistability for $NG = 0.49 \mu\text{S}/\text{cm}^2$. The network can be either quiescent or in a stable sustained active state in a range of external inputs, $J^{\text{inp}} f^{\text{inp}}$, less than the threshold current, I_c .

are the same. This is because we assume that the interaction profiles depend solely on the identity (excitatory or inhibitory) of the *presynaptic* neurons, and that the single neuron properties of the two types of cells are the same. We denote the rate of the (*e* or *in*) neurons with PO θ and a stimulus orientation \rightarrow as $f(\theta - \rightarrow)$. These rates obey

$$f(\theta - \rightarrow) = \beta \left[\int_{-\pi/2}^{+\pi/2} \frac{d\theta'}{\pi} J(\theta - \theta') f(\theta' - \theta) + J_{LGN} f_{LGN}(\theta - \rightarrow) - T \right]_+ \quad (2.29)$$

where we replaced the sum over the synaptic recurrent inputs by an integration over the variable θ' which is a valid approximation for a large network. The recurrent interaction profile, $J(\theta)$, combines the effect of the excitatory and the inhibitory cortical inputs and has the form

$$J(\theta - \rightarrow) = \sum_{\alpha=e,in} \frac{J_\alpha}{\lambda_\alpha} \exp(-|\theta - \theta'|/\lambda_\alpha) \quad (2.30)$$

where $J_\alpha = N_\alpha \bar{G}_\alpha \tau_\alpha (E_\alpha - E_L - V_c)$, $\alpha = e, in$, and $J_{LGN} = \bar{G}_{LGN} \tau_e (E_e - E_L - V_c)$; τ_α denotes the excitatory and inhibitory synaptic time constants, N_e, N_{in} are the number of neurons in the excitatory and inhibitory populations, respectively. Equations 2.29-2.30 correspond to the rate equations 2.20-2.22 with the synaptic conductances and input firing rate which are given by Eqs. 2.11-2.12. In Appendix B we outline the analytical solution of Eqs. 2.29-2.30, which allows us to compute the neuronal activity and the synaptic conductances as functions of the model parameters.

We used the analytical solution of these rate equations to explore how the spatial pattern of activity of the hypercolumn depends on the parameters of the recurrent interactions (J_e, J_{in}, λ_e , and λ_{in}) and the stimulus properties.

2.6.1 Emergence of a ring attractor

We first consider the case of an untuned LGN input, i.e. $\epsilon = 0$. In this case, Eq. 2.29 has a trivial solution in which all the neurons respond at the same firing rate. As shown in Appendix B this solution is unstable when the spatial modulation of the effective interaction, Eq. 2.30, is sufficiently large. The condition for the onset of this instability is given by Eq. 2.41 in Appendix B. When the homogeneous solution is unstable, the system settles into a heterogeneous solution, which has the form $f(\theta) = M(\theta - \psi)$. The angle ψ is arbitrary and reflects the fact that the system is spatially invariant. The manifold of stable states that emerges in this system and break its spatial symmetry

is known as a ring attractor. The function M which represents the shape of the activity profile in each of these states can be computed analytically as described in Appendix B. Depending on which mode is unstable, the heterogeneous profile of activity which emerges consists of a single “hill” of activity or several such “hills”. In Appendix B we describe how the function M can be computed in the case of a state with a single “hill”.

As an example we consider the case $\lambda_e = 11.5^\circ$, $\lambda_{in} = 43^\circ$ and $\beta J_{in} = 0.73$. For this choice of parameters, Eq. 2.41 predicts that the state with a homogeneous response is stable for $\beta J_e < 0.87$ and unstable for $\beta J_e > 0.87$. At $\beta J_e = 0.87$ the unstable mode corresponds to the first Fourier mode. Therefore, the instability at this point should give rise to a heterogeneous response with a unimodal profile of activity (a single ‘hill’). This is confirmed by the numerical solution of the rate equations, Eqs. 2.29-2.30.

Using the mapping prescriptions, Eq. 2.21, these results can be translated into the prediction that if $\lambda_e = 11.5^\circ$, $\lambda_{in} = 43^\circ$, $N_{in}\bar{G}_{in} = 0.333 \text{ mS/cm}^2$, the homogeneous state is stable for conductance-based model if $N_e\bar{G}_e < 0.138 \text{ mS/cm}^2$, but that it is unstable for $N_e\bar{G}_e > 0.138 \text{ mS/cm}^2$. We tested whether these predictions coincide with the actual behavior of the conductance-based model in numerical simulations. Figure 2.4A and B shows raster plots of the network for $N_e\bar{G}_e = 0.133 \text{ mS/cm}^2$ and $N_e\bar{G}_e = 0.143 \text{ mS/cm}^2$, respectively. For $N_e\bar{G}_e = 0.133 \text{ mS/cm}^2$ neurons in all the columns responded in a similar way. This corresponds to the homogeneous state of the rate model. Moreover, in this simulation the average population firing rate was $f = 18$ spikes/sec, in excellent agreement with the value predicted from the rate model for the corresponding parameters ($f = 18.05$ spikes/sec). In contrast, for $N_e\bar{G}_e = 0.143 \text{ mS/cm}^2$ the network does not respond homogeneously to the stimulus. Instead, a unimodal hill of activity appears. This is congruent with the prediction of the rate model. Since the external input is homogeneous, the location of the peak is arbitrary. In the numerical simulations the activity profile slowly moves due to the noise in the LGN input.

The stability analysis of the homogeneous state of the rate model shows that if $\beta J_{in} > 0.965$, which corresponds to $N_{in}\bar{G}_{in} = 0.443 \text{ mS/cm}^2$ the mode $n = 2$ is the one which first becomes unstable when J_e increases. This suggests that in this case the profile of activity which emerges through the instability is bimodal. Numerical simulations of the full conductance-based model were

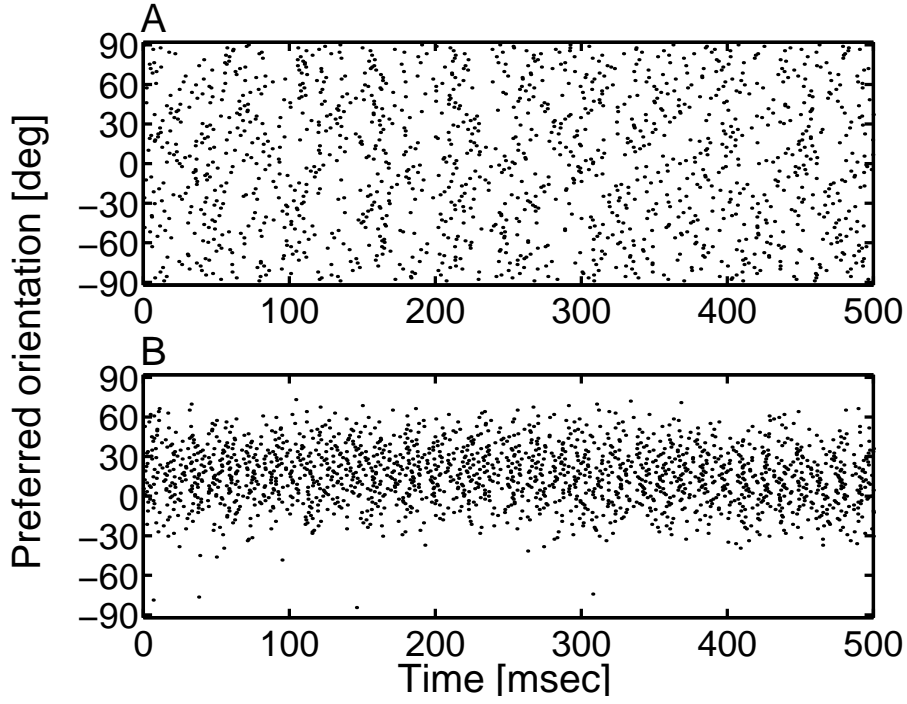


Figure 2.4: Symmetry breaking leading to a unimodal activity profile. A network with $N_e = N_{in} = 1600$ was simulated for two values of the maximal conductance of the excitatory synapses. The external input to the network is homogeneous ($\epsilon = 0$). The input rate is $\bar{f}_{LGN} = 2700$ Hz. The maximal conductance of the input synapses is $G^{inp} = 0.0025$ mS/cm². Parameters of the interactions are: $\lambda_e = 11.5^\circ$, $\lambda_{in} = 43^\circ$, $N_{in}\bar{G}_{in} = 0.333$ mS/cm². The time constants of the synapses are: $\tau_e = \tau_{in} = 3$ msec. The analytical solution of the rate model equations predicts that for $N_e\bar{G}_e < 0.138$ mS/cm² the response of the network to the input is homogeneous and that for $N_e\bar{G}_e > 0.138$ mS/cm² it is unimodal. A: Raster plot of the network for $N_e\bar{G}_e = 0.133$ mS/cm² showing that the response is homogeneous. B: $N_e\bar{G}_e = 0.143$ mS/cm² showing that the response is a unimodal hill of activity. The noise which is present in the system induces a slow random wandering of the hill of activity.

found to be in excellent agreement with this expectation (see Figure 2.5).

2.6.2 Tuning of firing rates and synaptic conductances

We consider now the case of a tuned LGN input, which corresponds to $\epsilon > 0$. Eq. 2.29 shows that in general the response of a neuron with PO θ depends on the stimulus orientation θ_0 through the difference $\theta - \theta_0$, namely that: $f(\theta, \theta_0) = M(|\theta - \theta_0|)$. For fixed θ_0 , when θ varies $f(\theta, \theta_0)$ is the profile of activity of the network in response to a stimulus of orientation θ_0 . Conversely, when θ is fixed and θ_0 varies, $f(\theta, \theta_0)$ is the tuning curve of the neuron with PO θ . Therefore, the function M determines the tuning curve of the neurons in the model.

We now compare the tuning curves of the neurons computed in the framework of the rate model with those in the corresponding simulations of the conductance-based network. Specifically, we take $\epsilon = 0.175$ and $\bar{f}_{LGN} = 3400$ Hz. For these values, in the absence of recurrent interactions, the response of the neurons to the input exhibits broad tuning. This is shown in Figure 2.6 (dashed line). The recurrent excitation can sharpen the tuning curves and also amplify the neuron response as shown in Figure 2.6. In this figure we plotted the neuronal tuning curve when the parameters of the interactions are $\lambda_e = 6.8^\circ$ and $\lambda_{in} = 43^\circ$, $N_e \bar{G}_e = 0.125$ mS/cm², $N_{in} \bar{G}_{in} = 0.333$ mS/cm². The solid line was computed from the solution of the mean-field equations of the corresponding rate model. This solution indicates that the tuning width is $\theta_C = 30^\circ$ and that the maximal firing rate is $f_{max} = 75.5$ spikes/sec. The circles are from the numerical simulations of the conductance-based model. The agreement with the analytical predictions from the rate model is very good.

The input conductances of the neurons in V1 change upon presentation of a visual stimulus. Experimental results (Borg-Graham, Monier, & Fregnac, 1998; Carandini, Anderson, & Ferster, 2000) indicate that with large stimulus contrasts, these changes have typical values of 60% when the stimulus is presented at null orientation whereas they can be as large as 250%-300% at optimal orientation. We applied our approach to study the dependence of the total change in input conductance on the space constants of the interactions for a given LGN input. We assume that the cortical circuitry sharpens and amplifies the response such that the tuning curves have a given width θ_C and amplitude f_{max} . We compute the interaction parameters to achieve tuning curves with a given width, θ_C , and a given maximal discharge rate, f_{max} . To be specific, we take

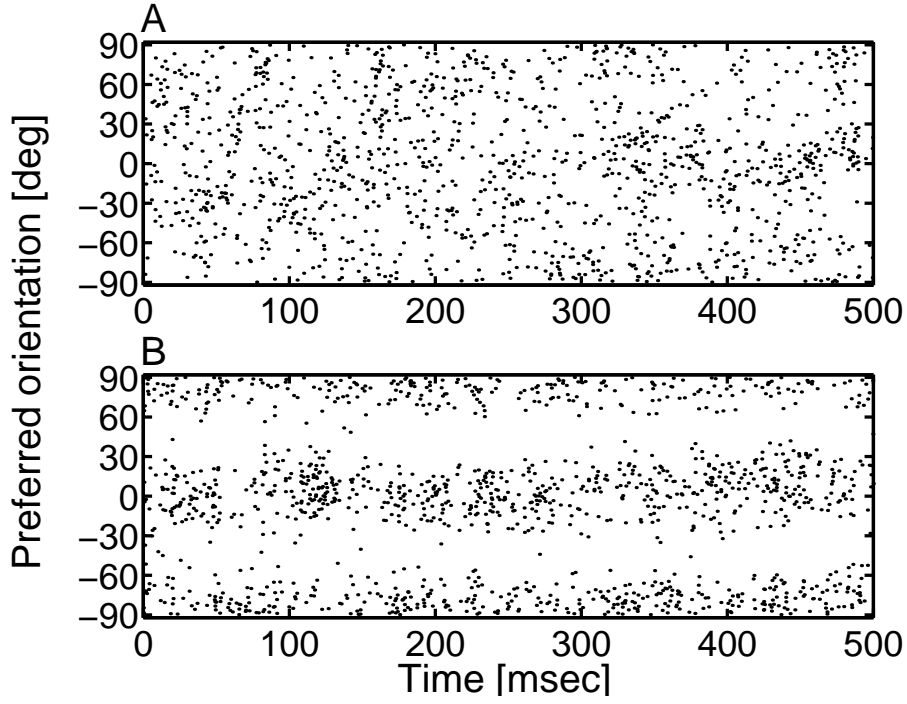


Figure 2.5: Symmetry breaking leading to a bimodal activity profile. The size of the simulated network is $N_e = N_{in} = 1600$. The input rate is $\bar{f}_{LGN} = 2700$ Hz. The maximal conductance of the input synapses is $G^{imp} = 0.0025$ mS/cm^2 . The parameters of the interactions are: $\lambda_e = 11.5^\circ$, $\lambda_{in} = 43^\circ$, $N_{in}G_{in} = 1.33$ mS/cm^2 . The time constants of the synapses are: $\tau_e = \tau_{in} = 3$ msec. The analytical solution of the rate model equations predicts that for $N_e\bar{G}_e < 0.196$ mS/cm^2 the response of the network to the input is homogeneous and that for $N_e\bar{G}_e > 0.196$ mS/cm^2 it is bimodal. A: Raster plot of the network for $N_e\bar{G}_e = 0.19$ mS/cm^2 showing that the response is homogeneous. The average firing rate in the network in the simulation is $f = 3.2$ spikes/sec which is in good agreement with the prediction of the rate model ($f = 2.9$ spikes/sec) B: $N_e\bar{G}_e = 0.138$ mS/cm^2 showing that the response is bimodal. The noise which is present in the system induces a slow random wandering of the pattern of activity.

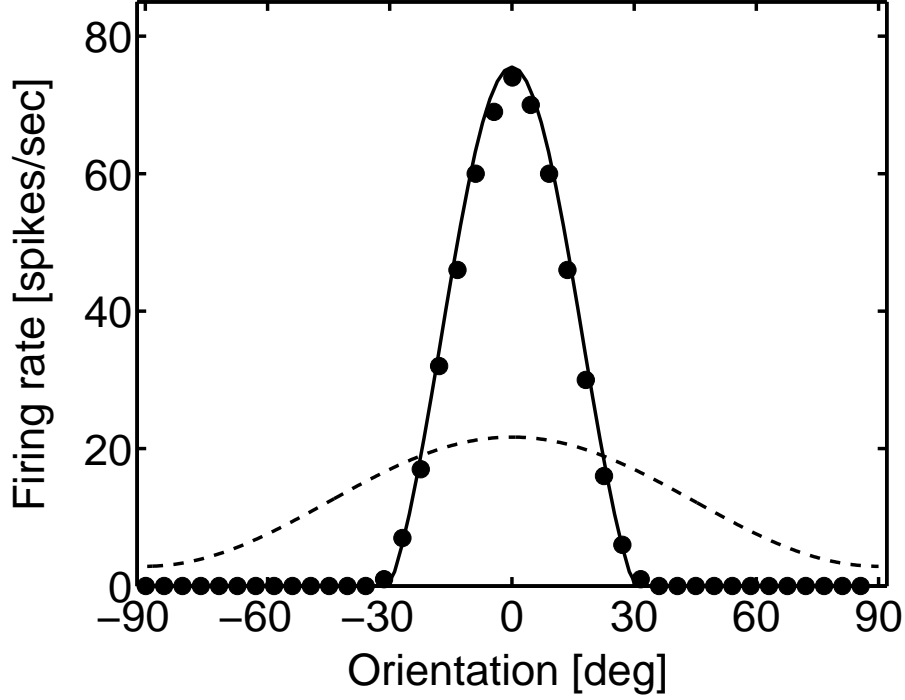


Figure 2.6: Tuning curves of the LGN input (dashed line) oriented at $\theta_0 = 0^\circ$ and the spike response of a neuron with preferred orientation $\theta = 0^\circ$ (solid line and circles). The LGN input parameters are $\epsilon = 0.175$, $g_{\text{inp}} = 0.0025 \text{ mS/cm}^2$, $\bar{f}_{LGN} = 3400 \text{ Hz}$. The interaction parameters are: $\lambda_e = 6.3^\circ$, $\lambda_{in} = 43^\circ$, $N_e \bar{G}_e = 0.125 \text{ mS/cm}^2$, $N_{in} \bar{G}_{in} = 0.467 \text{ mS/cm}^2$. The time constant of the synapses are: $\tau_e = \tau_{in} = 3 \text{ msec}$. The circles are from numerical simulations with $N_e = N_{in} = 1600$ neurons. The response of the neuron was averaged over 1 sec of simulation. The solid line was obtained by solving the rate model with the corresponding parameters.

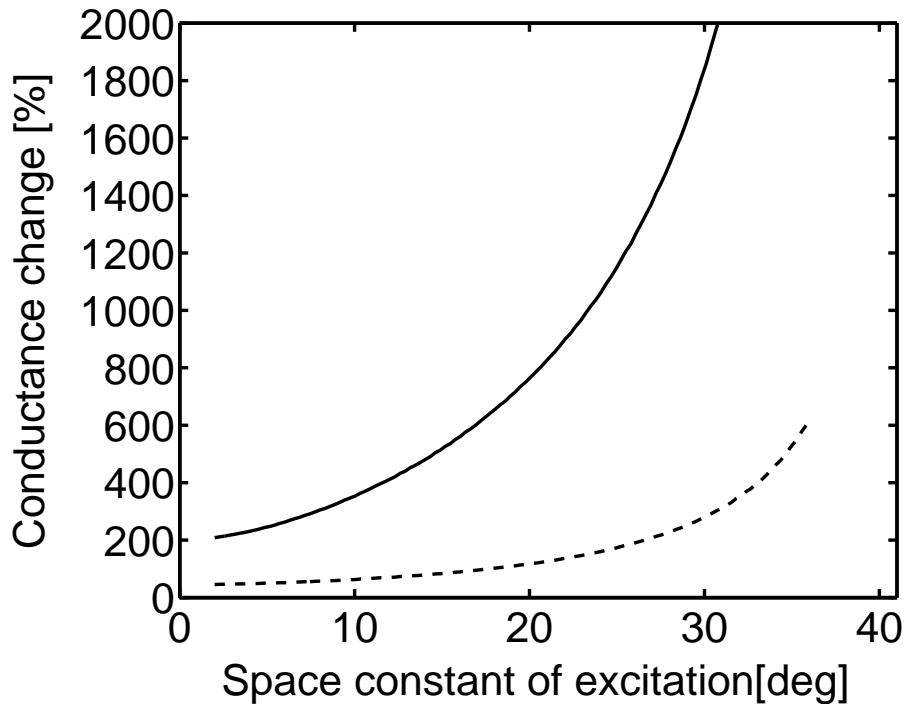


Figure 2.7: The change in the input conductance in iso (solid line) and cross-orientation (dashed line). The lines were computed as explained in the text.

$\theta_C = 30.5^\circ$, $f_{max} = 70$ spikes/sec. We also fixed the space constant of the inhibitory interaction, $\lambda_{in} = 43^\circ$, and varied the value of λ_e . For each value of λ_e we evaluated the values of J_e and J_{in} that yield the desired values of θ_C and f_{max} . Subsequently, for each set of parameters, we computed the changes in the input conductance of the neurons, relative to the leak conductance, for a stimulus presented in iso and cross orientation. These changes, denoted by Δg_{iso} and Δg_{cross} respectively, are increasing functions of λ_e , as shown in Figure 2.7. Actually, it can be shown analytically from the mean field equations of the rate model that the conductance changes diverge when $\lambda_e \rightarrow \lambda_{in}$. This is because in that limit the net interaction is purely excitatory with an amplitude which is above the symmetry breaking instability.

The rate model also allows us to estimate the separate contributions of the LGN input, the recurrent excitation and the cortical inhibition to the change in total input conductance induced by the LGN input. An example of the tuning curves of these contributions is shown in Figure 2.8 where they are compared with the results from the simulations of the full model. These results

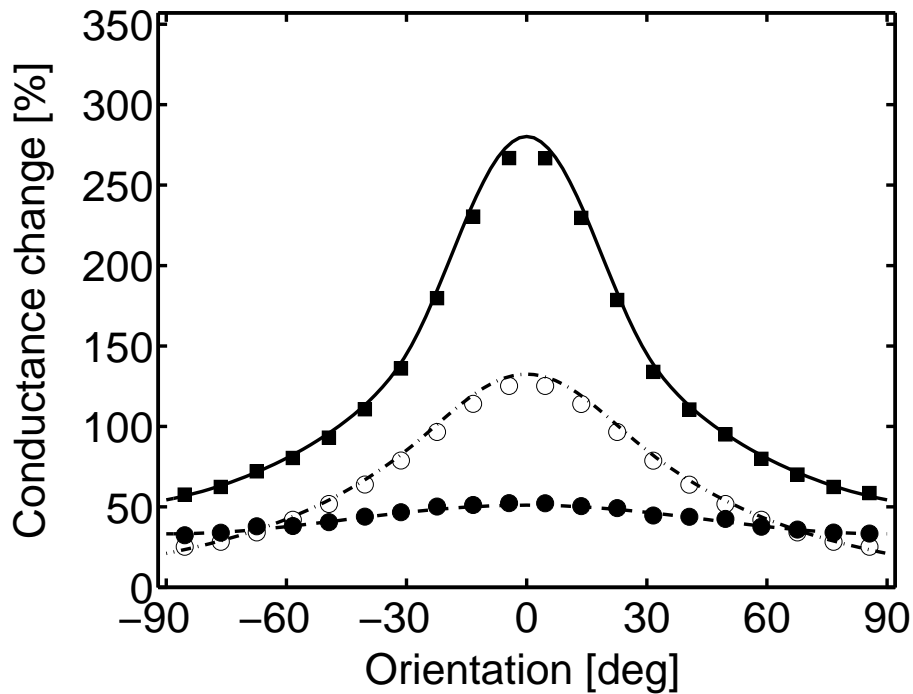


Figure 2.8: Tuning curves of conductances. Solid line: total change. Dotted line: the contribution to the total change from the LGN input. Dashed line: The contribution to the total change from the recurrent excitatory interactions. Dash-dotted line: the contribution to the total change from the recurrent inhibitory interactions. These lines were obtained from the solution of the mean-field equations of the rate model. The parameters are as in figure 2.6. The squares, circles, triangles and diamonds are from numerical simulations. Same parameters as in figure 2.6.

indicate that for the chosen parameters, most of the input conductance contributed by the recurrent interactions comes from the cortical inhibition. This is despite the fact that the spike discharge rate is greatly amplified by the cortical excitation.

2.7 Rate Response of a Single Neuron to a Time-dependent Input

The analysis of the previous sections focused on situations in which the firing rates of the neurons were approximately constant in time. We now turn to the question of firing rate dynamics, namely how to describe the neuronal firing response in the general situation in which the firing rates are

time-dependent.

We first study the response of a single neuron to a noisy sinusoidal input, Eq. 2.10. The firing rate of the neuron (averaged over the noise) can be expanded in a Fourier series:

$$f(t) = f_0 + f_1 \cos(\omega t + \phi) + \dots \quad (2.31)$$

where f_0 is the mean firing rate and f_1 and ϕ are the amplitude and phase of the first harmonic (the Fourier component at the stimulation frequency). Here, we only consider cases in which the modulation of the external input is not overly large, so that there is no rectification of the firing rate by the threshold.

Under these conditions, our simulations show that harmonics with orders higher than 1 are negligible (results not shown). Therefore, the response of the neuron can be characterized by the average firing rate f_0 and the modulation f_1 , $f_1 < f_0$. Our simulations also show that the mean response, f_0 , depends only weakly on the modulation frequency of the stimulus or on the stimulus amplitude, and can be well described by the steady-state frequency-current relation. This is shown in Figure 2.9 (left panels), where the predictions from the rate model (solid horizontal lines) are compared with the mean output rates in the numerical simulations of the conductance-based model (hollow circles).

Figure 2.9 also shows the amplitude of the first harmonic of the response and its phase shift ϕ as a function of the modulation frequency $\nu \equiv \omega/2\pi$ (filled circles). It should be noted that the raw response of the neuron reflects two filtering processes of the external input rate: the low-pass filtering induced by the synaptic dynamics, and the filtering induced by the intrinsic dynamics of the neuron and its spiking mechanism. To better elucidate the transfer properties of the neuron, we remove the effect of the synaptic low-pass filtering on the amplitude and phase of the response. (This corresponds to multiplying f_1 by $\sqrt{1 + \omega^2(\tau^{\text{inp}})^2}$ and subtracting $\tan^{-1}(-\omega\tau^{\text{inp}})$ from the phase). The results for two values of the mean output rate, $f_0 \simeq 30 \text{ spikes/sec}$ and $f_0 \simeq 60 \text{ spikes/sec}$, and for two values of the input noise coefficient of variation, $\Delta = 0.3$ ($f_0^{\text{inp}} = 1125 \text{ Hz}$) and $\Delta = 0.18$ ($f_0^{\text{inp}} = 3125 \text{ Hz}$), are presented. Clearly, both the amplitude and the phase of the response depend on the modulation frequency. Of particular interest is the fact that the amplitude exhibits a resonant behavior for modulation frequencies close to the mean firing rate of the neuron, f_0 . The main effect of increasing the coefficient of variation of the noise is to broaden the resonance

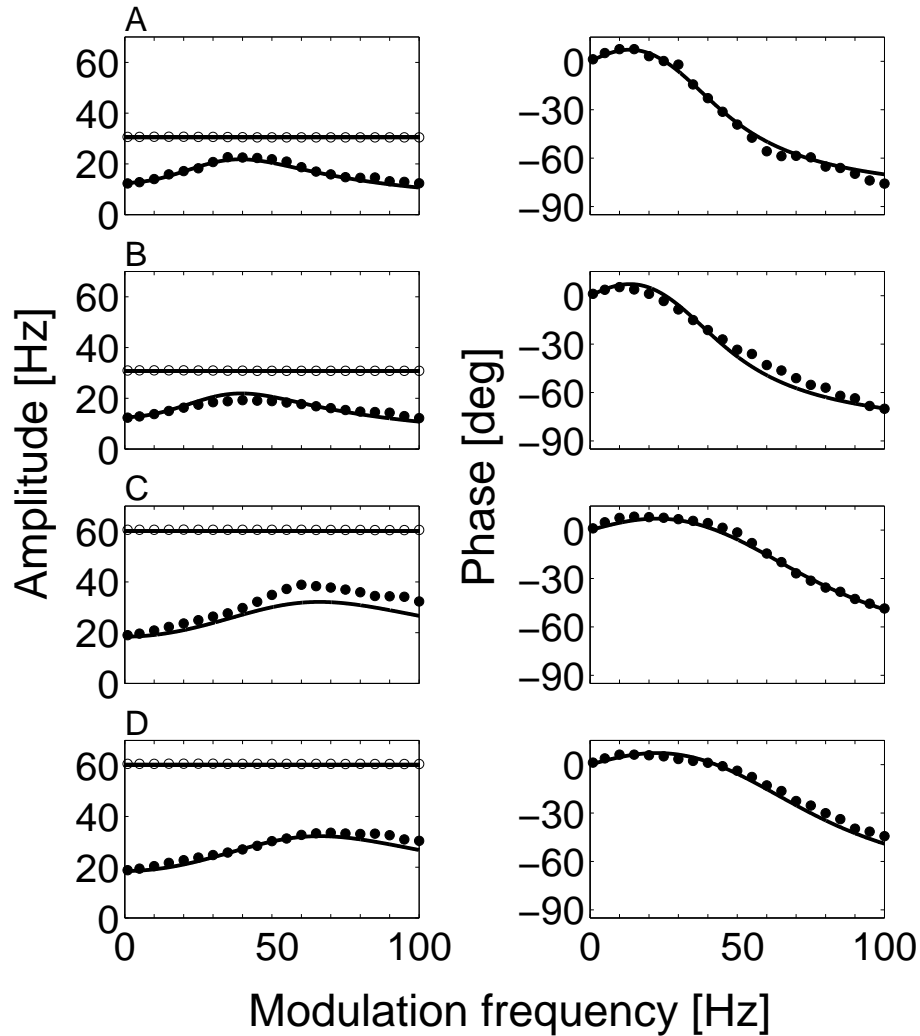


Figure 2.9: Mean, amplitude and phase of single neuron rate response as a function of the modulation frequency. The left panels show the mean output rate of the neuron (hollow circles) and the amplitude of the response (filled circles). The right panels show the phase of the response. The solid curves are the predictions of the dynamic rate model (see text). The external input was designed to produce a mean output rate of about ~ 30 spikes/sec in A and B and ~ 60 spikes/sec in C and D. A and C show the responses to inputs with a small noise coefficient of variation, $\Delta = 1/\sqrt{2f_0^{\text{inp}}\tau^{\text{inp}}} = 0.18$, while B and D show the responses to inputs with a high noise level, $\Delta = 0.3$.

peak.

As in our analysis of the steady-state properties, the external synaptic input can be decomposed here into a current term and a conductance term which are now time-dependent. The simplest dynamic model would be to assume that the same f-I relation that was found under steady-state conditions, Eqs. 2.2-2.3, holds when the applied current and the passive conductance are time dependent. In our case, this would take the form $f(t) = \beta[I(t) - (I_c^o + V_c g(t))]_+$. This model predicts that the response amplitude does not depend on the modulation frequency and that the phase shift is always zero, in contrast to the behavior of the conductance-based neuron (see Figure 2.9). To account for the dependence of the modulation frequency we extend the model by assuming that it has the form:

$$f(t) = \beta[I_{\text{filt}}(t) - (I_c^o + V_c g_{\text{filt}}(t))] \quad (2.32)$$

where $I_{\text{filt}}(t)$ and $g_{\text{filt}}(t)$ are filtered versions of the current and conductance terms, respectively. For simplicity we use the same filter for both current and conductance. A first-order linear filter can be either a high-pass or a low-pass filter. Since the dependence of the response amplitude on the modulation frequency has a band-pass nature, a first-order linear filter would not be suitable. Thus, we assume a second-order linear filter. The filter for the current is described by

$$\frac{1}{\omega_0^2} \frac{d^2 I_{\text{filt}}}{dt^2} + \frac{1}{\omega_0 Q} \frac{dI_{\text{filt}}}{dt} + I_{\text{filt}} = I + \frac{a}{\omega_0} \frac{dI}{dt} \quad (2.33)$$

and a similar equation (with the same parameters) defines the conductance filter. This is the equation of a damped harmonic oscillator with a resonance frequency $\omega_0/2\pi$ and 'Q-factor' Q (Q is defined as ω_0 over the width at half height of the resonance curve). Note that the driving term is a linear combination of the driving current input and its derivative.

We investigated the behavior of the optimal filter parameters over a range of mean output rates, 10 – 70 spikes/sec, and a range of input noise levels, $\Delta = 0.15 - 0.3$. For lower noise levels the response profile contains additional resonant peaks (both subharmonics and harmonics), that cannot be accounted for by the linear filter of Eq. 2.33. For each mean output rate and input noise level we ran a set of simulations with modulation frequencies ranging from 1 Hz to 100 Hz, and then numerically found the set of filter parameters that gave the best fit for both amplitude and phase of the response. The variation of the optimal values for Q and a was small under the range of input

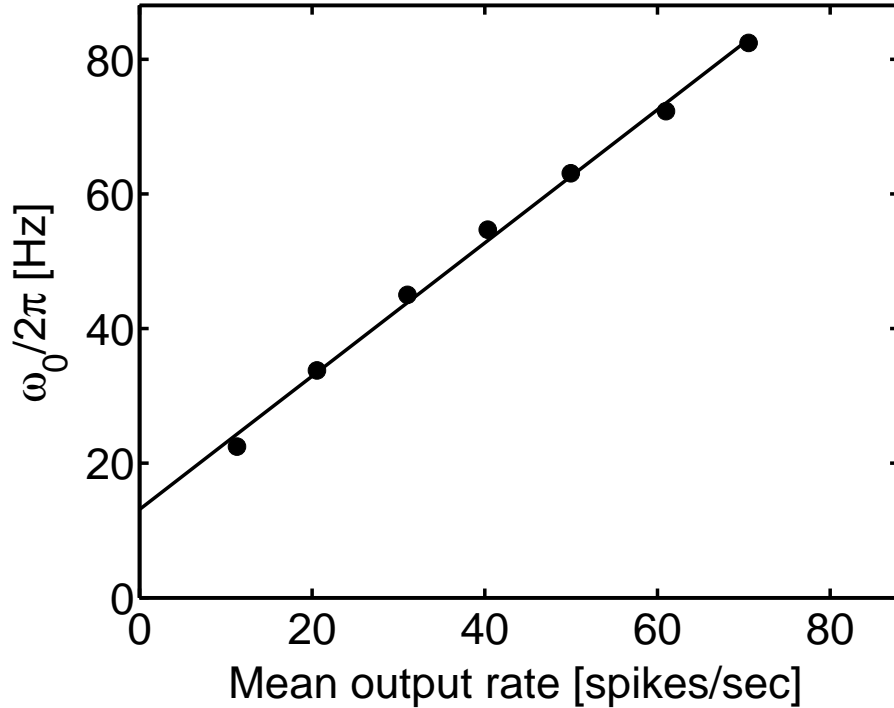


Figure 2.10: Resonance frequency ($\omega_0/2\pi$) as a function of the mean output rate. For each value of the mean output rate, optimal values of ω_0 were averaged over 4 noise levels: $\Delta = 0.15, 0.18, 0.22, 0.3$. The range of modulation frequencies for the fit was 1 – 100 Hz . The optimal linear fit (solid line) has a slope of 1 and an offset of 13.1.

parameters we considered, with mean values of $Q = 0.85$ and $a = 1.9$. The resonance frequency $\omega_0/2\pi$ depends only weakly on the noise level but increases linearly with the mean output rate of the neuron, with a slope of 1, $\omega_0/2\pi = f_0 + 13.1$, as shown in Figure 2.10. The solid curves in Figure 2.9 show the predictions of Eq. 2.33 for the amplitude and phase of the rate response using the mean values of Q and a and the linear fit for ω_0 , mentioned above. The results show that this model is a reasonable approximation of the response modulation of the conductance based neuron.

So far we have dealt with the firing rate response of a single neuron to input composed of a single sinusoid. We also tested the response of the neuron to a general broadband stimulus. For this, we used a Poissonian input characterized by a mean rate and a time varying envelope that consisted of a superposition of 10 cosine functions with random frequencies, amplitudes and phases. The mean input rate was chosen to obtain a mean output firing rate ~ 30 spikes/sec.

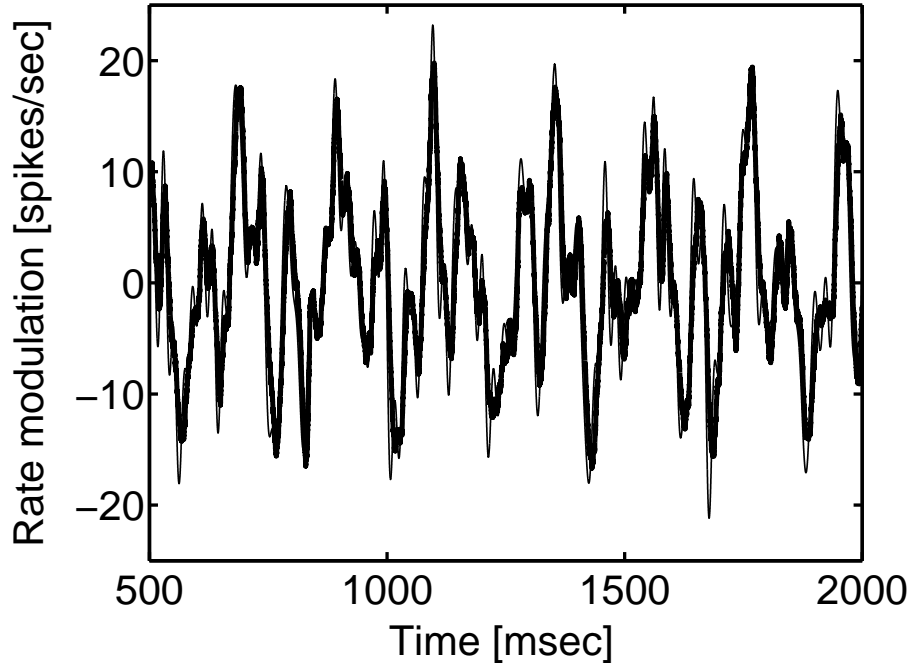


Figure 2.11: Single neuron response to a broadband signal. The thick curve shows the rate of the conductance based neuron while the thin curve shows the prediction of the rate model. The input consisted of a superposition of 10 cosine functions with random frequencies, amplitudes and phases (see text for specifics). For clarity, only the fluctuations around the mean firing rate (~ 30 spikes/sec) are shown.

The modulation frequencies were drawn from a uniform distribution between 0 and 60Hz and the phases from a uniform distribution between 0 and 360° . The relative amplitudes were drawn from a uniform distribution between 0 and 1 and to avoid rectification they were normalized so that their sum is 0.8. The results are presented in Figure 2.11. They show that the response to a broadband stimulus can be predicted from the response to each of the Fourier components. This confirms the validity of our linearity assumption, Eq. 2.32.

2.8 Response of a Neuronal Population to a Time Periodic Input

We now apply the approach of the previous section to study the dynamics of a network of interacting neurons in response to a time dependent input. Here we consider the case of a large homogeneous

network of N fully connected neurons which receive an external noisy oscillating input. The input to each neuron is a Poisson process with a sinusoidal envelope. In addition, the firing times of the inputs that converge on different cells in the network are uncorrelated, so that the Poissonian fluctuations in the inputs to the different cells are uncorrelated. Nevertheless, since the underlying oscillatory rate of the different inputs is coherent the network response will have a synchronized component. To construct a simple model of the dynamics of the network rate we first define two *conductance rate variables*: $r(t) = g(t)/(NG\tau_e)$ and $r^{\text{inp}}(t) = g^{\text{inp}}(t)/(G^{\text{inp}}\tau_e)$ where G and G^{inp} are the peak conductances of the recurrent and afferent synapses, respectively.

Averaging Eq. 2.7 over all the neurons in the network yields the following equation for the population conductance rate,

$$\tau_e \frac{dr}{dt} = -r + f(t) \quad (2.34)$$

where $f(t)$ is the instantaneous firing rate of the network per neuron. A similar equation holds for $r^{\text{inp}}(t)$, which is a low-pass filter of $f^{\text{inp}}(t)$. The equation for $f(t)$ is

$$f(t) = \beta [J^{\text{inp}}\rho^{\text{inp}} + J\rho - I_c] \quad (2.35)$$

where J^{inp} and J are given in Eq. 2.28. The quantities $\rho(t)$ and $\rho^{\text{inp}}(t)$ are obtained from $r(t)$ and $r^{\text{inp}}(t)$ respectively using the filter in Eq. 2.33, i.e.

$$\frac{1}{\omega_0^2} \frac{d^2\rho}{dt^2} + \frac{1}{\omega_0 Q} \frac{d\rho}{dt} + \rho = r + \frac{a}{\omega_0} \frac{dr}{dt} \quad (2.36)$$

and a similar equation holds for ρ^{inp} . This yields a set of self-consistent equations which determine the firing rate of the neurons in the network, $f(t)$. For response modulation amplitudes smaller than the mean firing rate, the dynamics are linear and can be solved analytically (see Appendix C). Figure 2.12 shows the mean, amplitude and phase of the network response $f(t)$ obtained by simulating the dynamics of the conductance-based network together with the analytical predictions of the rate model. (As in the previous section, the filtering done by the input synapses was removed for purposes of presentation). The mean rate of the Poisson input and the strength of the synaptic interaction were chosen such that the mean firing rate of the network will be around 50 spikes/sec and the noise level will be $\Delta = 0.22$.

The results of Figure 2.12 reveal the effect of the recurrent interactions on the modulation of the network rate response. Qualitatively, the peak of the response amplitude profile is suppressed

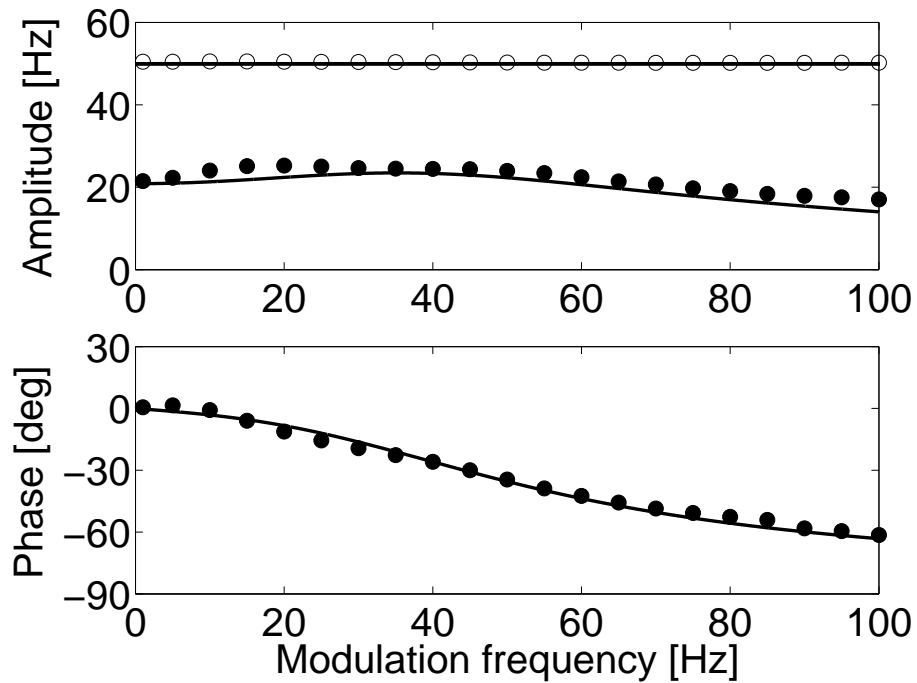


Figure 2.12: Response of an excitatory network to a sinusoidal stimulus. A: Mean (hollow circles) and amplitude (filled circles) B: phase. The network consists of $N = 4000$ neurons. The neurons are coupled all-to-all. The synaptic conductance is $NG = 0.039 \text{ mS/cm}^2$ and the synaptic time constant is $\tau_e = 5 \text{ msec}$. The reversal potential of the synapses was $V_e = 0 \text{ mV}$.

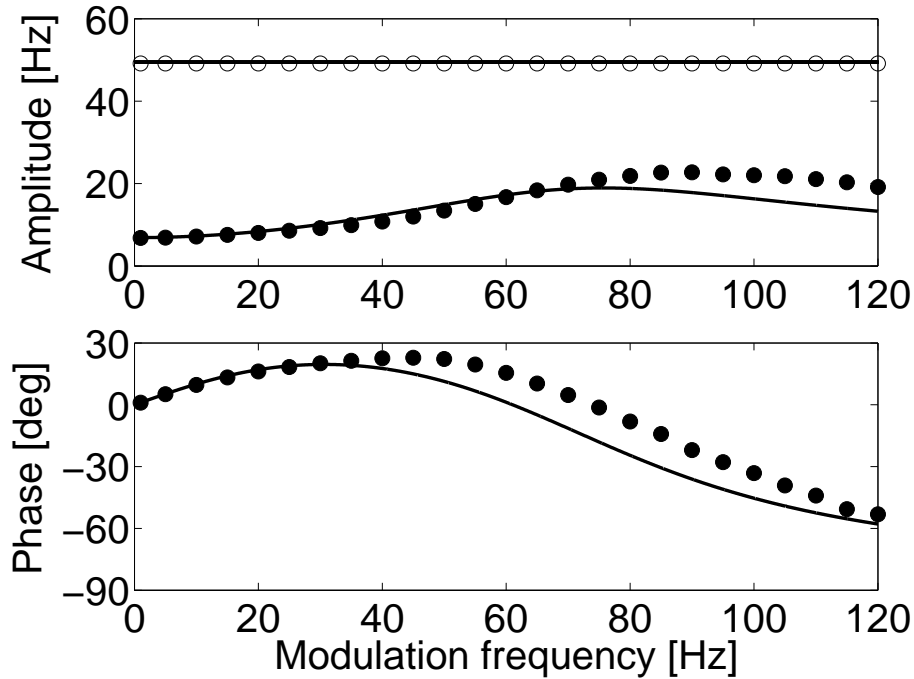


Figure 2.13: Response of an inhibitory network to a sinusoidal stimulus. A: Mean (hollow circles) and amplitude (filled circles) B: phase. The network consists of $N = 4000$ neurons. The neurons are coupled all-to-all. The synaptic conductance is $NG = 0.18 \text{ mS/cm}^2$ and the synaptic time constant is $\tau_e = 5 \text{ msec}$. The reversal potential of the synapses was $V_{in} = -80 \text{ mV}$.

and shifts to smaller frequencies due to the excitatory connectivity. In contrast, for an inhibitory network the model predicts that the peak response will be shifted to the right and that the resonant behavior will be more pronounced. This can be proved from Eq. 2.52 (Appendix C) by taking negative J . To test this prediction, we ran simulations of a uniform fully-connected network with inhibitory synapses (the reversal potential was -80 mV). The input parameters and the strength of the synaptic interaction were chosen to produce a mean firing rate around 50 spikes/sec and a noise level $\Delta = 0.3$. Shown in Figure 2.13 are the results of the numerical simulations together with the prediction of the rate model. These results provide additional strong support for our phenomenological time dependent rate response model.

2.9 Discussion

The rate models derived here are based on specific assumptions about single cell properties. The most important assumptions are: 1) the independence of the gain of the f-I curve of the leak conductance, g_L (Eq. 2.2), and 2) the approximated linear dependence of the threshold current on g_L (Eq. 2.3). This means that shunting conductances have a subtractive effect rather than a divisive one. This is in agreement with previous modeling studies (Holt and Koch, 1997) and with the properties of the conductance-based model neuron used in our study (Figure 2.2). Recent experiments using the dynamic clamp technique provide additional support for this assumption (Brizzi, Hansel, Meunier, van Vreeswijk, & Zytnicki, 2001; Chance, Abbott & Reyes, 2002). A further simplifying assumption, supported by experiments, is that in a broad range of input currents and output firing rates the f-I curves of cortical neurons can be well approximated by a threshold linear function. We used such a form of the f-I curve to show that the response of a single neuron to a stationary synaptic input can be well described by simple rate models with threshold nonlinearity. We applied our single neuron model to the study of a network of neurons, receiving an input generated by a set of synapses activated by uncorrelated trains of spikes with stationary Poisson statistics. In this case, we derived a rate model under the additional assumptions that the network is highly connected and that it is in an asynchronous state.

The mapping of the conductance-based model onto a rate model described in this work provides a correspondence between the “synaptic efficacy” used in rate models and biophysical parameters of the neurons. In particular, the sign of the synaptic efficacy is determined by the value of the reversal potential relative to $E_L + V_c$, where V_c is the threshold gain-potential of the post-synaptic neuron (Eq. 2.3). Furthermore, our theory enables the use of rate models to calculate synaptic conductances that are generated by the recurrent activity, allowing a quantitative comparison of predictions from rate models and conductance measurements in *in-vitro* and *in-vivo* intra-cellular experiments.

In the case of a fully connected network of excitatory neurons we showed that the firing rate of the neurons predicted by the rate model was highly congruent with simulation results in a broad range of synaptic conductances. This indicates that our rate model provides reliable results over a broad range of firing rates and conductance changes. Even for very high rates, incorporating a weak

quadratic nonlinearity is enough to account for the saturation of the neurons' firing rates. We also showed that our approach can be applied to networks with more complicated architectures, such as the conductance-based model of a hypercolumn in V1 analyzed in this work. The conditions for the stabilization of the homogeneous state can be correctly predicted from the mean-field analytical solution of the corresponding rate model. Furthermore, the profile of activity of the network and the tuning curve of the synaptic conductances can be calculated. Our results show that in order to obtain changes in input conductances which are similar to those found experimentally one has to assume that the space constant of the excitatory feedback is much smaller than the space constant of the inhibitory interactions. However this conclusion may depend on our assumption of interaction profiles which are identical for excitatory and inhibitory targets.

To extend our approach to the case of time dependent inputs, we studied the response of the single neuron to noisy input which is modulated sinusoidally in time. We showed that this response can be described by a rate model in which the neuron responds instantaneously to an effective input, which is a filtered version of the actual one. This is similar to the approach of Chance et al (2001), who studied the responses of spiking neurons to oscillating input currents. Our description is valid provided that the input is sufficiently noisy to broaden resonances and that its modulation is small compared to its average to avoid rate rectification. Interestingly, if these assumptions are satisfied, we found that the neuron essentially behaves like a linear device even if the modulation of the input is substantial. This allowed us to derive a rate model which provides a good description for the dynamics of a homogeneous network of neurons receiving a time dependent input.

Our derivation of rate equations for conductance-based models can be compared to the one suggested by Ermentrout (1994) (see also Rinzel & Frankel, 1992). Ermentrout derived a rate model in the limit where the synaptic time constants are much longer than the time constants of the currents involved in the generation of spikes, as well as in the inter-spike interval. In this case, the neuronal dynamics on the scale of the synaptic time constants can be reduced to rate equations of the type of Eq. 2.34 in which the rate variables, $r(t)$, represent the instantaneous activity of the slow synapses. However, in Ermentrout's approach, the firing rate $f(t)$ is given by an equation similar to Eq. 2.35 with $\rho^{\text{inp}} \equiv r^{\text{inp}}$ and $\rho \equiv r$. This is in contrast to our approach, where ρ^{inp} and ρ are filtered versions of r^{inp} and r . The two approaches become equivalent in the limit of slow

varying input and slow synapses, i.e., $\omega_0 \gg 1/t^{\text{inp}}, 1/\tau_s$ where t^{inp} is the typical time scale over which the external input varies and τ_s is the synaptic time constant of the faster synapses in the network. The parameter ω_0 is the resonance frequency.

Knight (Knight, 1972a, 1972b) studied the response of integrate-and-fire neurons to periodic input. He concluded that a resonant behavior occurred for input modulation frequencies around the average firing rate of the neurons. He also showed that white noise in the input broadens the resonance peak. Similar results were obtained by Gerstner in the framework of spike response models (Gerstner, 2000). This resonance has also been reported by Brunel et al. (2001) and Fourcaud and Brunel (2002), who studied analytically the firing response of a single integrate-and-fire neuron to a periodic input with temporally-correlated noise. Our results extend these conclusions to conductance-based neurons. An interesting finding of the present work is that the resonance frequency of the rate response increases linearly as a function of the mean output rate with a slope of one. It would be interesting to derive this relationship as well as the other parameters of our rate dynamics model from the underlying conductance-based dynamics. In addition, applications of our approach to other single-cell conductance-based models and to more complicated network architectures need to be explored. For instance, the effect of slow potassium adaptation currents should be taken into account. These currents are likely to contribute to the high-pass filtering of the neuron as shown by Carandini et al. (1996) in the case of an oscillating injected current.

In this work, we used an A-current as a mechanism for the linearization of the f-I curve. Other slow hyperpolarizing conductances might also achieve the same effect (Ermentrout, 1998). In particular, slow potassium currents, which are responsible for spike adaptation in many cortical cells, might be alternative candidates. However, linear f-I curves are also observed in cortical inhibitory cells, most of which do not show spike adaptation, but they probably do possess an A-current.

Lastly, we focused in this work on conductance-based network models of point neurons. An interesting open question is whether an appropriate rate model can also be derived for neurons with extended morphology. In conclusion, our results open up possibilities for applications of rate models to a range of problems in sensory and motor neuronal circuits in cortex as well as in other

structures with similar single neuron properties. Applying the rate model to a concrete neuronal system requires knowledge of a relatively small number of experimentally measurable parameters that characterize the f-I curves of the neurons in the system. In addition, the rate model requires knowledge of the gross features of the underlying connectivity pattern, as well as an estimate of the order of magnitude of the associated synaptic conductances.

Acknowledgments: We thank L. Abbott for fruitful discussions and Y. Loewenstein and C. van Vreeswijk for careful reading of the manuscript. This research was partially supported by a grant from the US-Israel Binational Science Foundation and a grant from the Israeli Ministry of Science and the French Ministry of Science and Technology. This research was also supported by the Israel Science Foundation (Center of Excellence Grant-8006/00).

2.10 Appendix A: Model Neuron

This appendix provides the specifics for the equations of the model neuron used here. The dynamics of our single neuron model consist of the following equations

$$C_m \frac{dV}{dt} = -I_L - I_{Na} - I_K - I_A + I^{\text{app}} \quad (2.37)$$

The leak current is given by $I_L = g_L(V - E_L)$. The sodium and the delayed rectifier currents are described in a standard way: $I_{Na} = \bar{g}_{Na} m^3 h (V - E_{Na})$ for the sodium current and $I_K = \bar{g}_K n^4 (V - E_K)$ for the delayed rectifier current. The gating variables $x = h, n$ satisfy the relaxation equations: $dx/dt = (x_\infty - x)/\tau_x$. The functions x_∞ , ($x = h, n, m$), and τ_x are: $x_\infty = \alpha_x/(\alpha_x + \beta_x)$, and $\tau_x = \phi/(\alpha_x + \beta_x)$ where $\alpha_m = -0.1(V+30)/(\exp(-0.1(V+30)) - 1)$, $\beta_m = 4 \exp(-(V+55)/18)$, $\alpha_h = 0.07 \exp(-(V+44)/20)$, $\beta_h = 1/(\exp(-0.1(V+14)) + 1)$, $\alpha_n = -0.01(V+34)/(\exp(-0.1(V+34)) - 1)$ and $\beta_n = 0.125 \exp(-(V+44)/80)$. We have taken: $\phi = 0.1$.

The A -current is: $I_A = \bar{g}_A a_\infty^3 b (V - E_K)$ with $a_\infty = 1/(\exp(-(V+50)/20) + 1)$. The function $b(t)$ satisfies $db/dt = (b_\infty - b)/\tau_A$ with: $b_\infty = 1/(\exp((V+80)/6) + 1)$. For the sake of simplicity, the time constant, τ_A , is voltage independent.

The other parameters of the model are: $C_m = 1 \mu F/cm^2$, $\bar{g}_{Na} = 100 mS/cm^2$, $\bar{g}_K = 40 mS/cm^2$. Unless specified otherwise, $g_L = 0.05 mS/cm^2$, $\bar{g}_A = 20 mS/cm^2$ and $\tau_A = 20 msec$. The reversal potentials of the ionic and synaptic currents are: $E_{Na} = 55 mV$, $E_K = -80 mV$, $E_L = -65 mV$, $E_e = 0 mV$ and $E_{in} = -80 mV$. The external current is I_{app} (in $\mu A/cm^2$).

2.11 Appendix B: Model of a hypercolumn in V_1 : the mean-field equations

The instability of the homogeneous state: For a homogeneous input a trivial solution to the fixed point equation, Eq. 2.29, corresponds to a state in which the responses of all the neurons are the same. However, this homogeneous state can be unstable if the spatial modulation of the interactions is sufficiently large. The condition for this instability can be derived by solving Eq. 2.29 in the limit of a weakly heterogeneous input $f_{LGN}(\theta)$ as follows.

The Fourier expansion of f_{LGN} is:

$$f_{LGN}(\theta) = \sum_{n=1}^{\infty} f_n \exp(2in\theta) \quad (2.38)$$

where the coefficients f_n are complex numbers. For weakly heterogeneous inputs, all the coefficients but f_0 are small. In the parameter regime where the homogeneous state is stable the response of the network to this input will be weakly heterogeneous. Therefore the Fourier expansion of the activity profile is:

$$m(\theta) = \sum_{n=1}^{\infty} m_n \exp(2in\theta) \quad (2.39)$$

where all the coefficients but m_0 are small. Substituting Eq. 2.38, 2.39 in Eq. 2.29 and computing the coefficients m_n , $n > 0$ perturbatively one finds:

$$m_n = \frac{J_{LGN} f_n}{1 - \beta J_n} \quad (2.40)$$

where J_n are the Fourier coefficients of the interactions.

The coefficient m_n diverges if $\beta J_n = 1$. This divergence indicates an instability of the homogeneous state. This instability induces a heterogeneous profile of activity with n peaks.

The coefficients J_n can be computed from Eq. 2.30. This yields the instability onset condition:

$$2\beta \left(J_e \frac{1 - (-1)^n \exp(-\pi/2\lambda_e)}{1 + 4n^2\lambda_e^2} + J_{in} \frac{1 - (-1)^n \exp(-\pi/2\lambda_{in})}{1 + 4n^2\lambda_{in}^2} \right) = 1 \quad (2.41)$$

Mean-field equations for profile of activity: We are interested in the case in which the LGN input is broadly tuned, $0 < \epsilon < 1/2$ and the effect of the interactions is sufficiently strong to sharpen substantially the response of the neurons. More specifically, we require that $f(\theta) = 0$ for all the neurons with POs such that $|\theta - \theta_0| > \pi/2$. In this case, the solution to Eqs. 2.29-2.30 can be found analytically. Taking $\theta_0 = 0$, without loss of generality, this solution has the form:

$$f(\theta) = A_0 + A_1 \cos(\mu_1\theta) + A_2 \sin(\mu_2\theta) + A_3 \cos(2\theta) \quad \text{for } |\theta| < \theta_c \quad (2.42)$$

$$f(t) = 0 \quad \text{for } \theta_c < |\theta| \quad (2.43)$$

The angle θ_c is determined by

$$f(\pm\theta_c) = 0 \quad (2.44)$$

Substituting Eq. (2.42) in Eq. (2.29) one finds that μ_1 and μ_2 are solutions (real or imaginary) of the equation:

$$\sum_{\alpha=E,I} \frac{2J_\alpha}{1 + \lambda_\alpha^2 x^2} = 1 \quad (2.45)$$

and that the coefficients A_0 and A_3 are given by:

$$A_0 = \frac{\bar{J}_{LGN} f_{LGN} (1 - \epsilon) - T}{1 - 2J_e - 2J_{in}} \quad (2.46)$$

$$A_3 = \frac{\bar{J}_{LGN} f_{LGN}}{1 - 2 \left(\frac{\mu_1 J_e}{1 + 4\mu_1^2} + \frac{\mu_2 J_{in}}{1 + 4\mu_2^2} \right)} \quad (2.47)$$

Finally, A_1 and A_2 are obtained from the two equations:

$$F(\lambda_\alpha, \mu, \mathbf{A}) = 0 \quad \alpha = E, I \quad (2.48)$$

where, $\mu = (\mu_1, \mu_2)$, $\mathbf{A} = (A_0, A_1, A_2, A_3)$ and F is the function

$$F(x, \mu, \mathbf{A}) = A_0 + \sum_{i=1,3} \frac{A_i}{1 + \mu_i x^2} (\cos(\mu_i \theta_c) - \mu_i x \sin(\mu_i \theta_c)) \quad (2.49)$$

with $\mu_3 = 2$. and the angle θ_c is determined by the condition:

$$A_0 + A_1 \cos(\mu_1 \theta_c) + A_2 \sin(\mu_2 \theta_c) + A_3 \cos(2\theta_c) = 0 \quad (2.50)$$

2.12 Appendix C: Rate dynamics in a neuronal population with uniform connectivity

By Fourier transforming Eqs. 2.34-2.36 (assuming that the firing rates are always positive) we obtain for $\omega \neq 0$

$$\hat{f}(\omega) = \beta(J^{\text{inp}} \hat{\rho}^{\text{inp}} + J \hat{\rho}) \quad (2.51)$$

where $\hat{\rho}(\omega) = B(\omega) \hat{r}(\omega)$, $\hat{r}(\omega) = L(\omega) \hat{f}(\omega)$, $\hat{\rho}^{\text{inp}}(\omega) = B(\omega) \hat{r}^{\text{inp}}(\omega)$, $\hat{r}^{\text{inp}}(\omega) = L(\omega) \hat{f}^{\text{inp}}(\omega)$. The low-pass filter $L(\omega)$ and band-pass filter $B(\omega)$ are given by $L(\omega) = 1/(1 + i\omega\tau_e)$ and $B(\omega) = (1 + ia\omega/\omega_0)/(1 + iw/\omega_0 - \omega^2/\omega_0^2)$.

Putting the above relations in Eq. 2.51 and solving for $\hat{f}(\omega)$ gives the transfer function of the network

$$\hat{f}(\omega) = \frac{\beta J^{\text{inp}} B(\omega) L(\omega)}{1 - \beta J B(\omega) L(\omega)} \hat{f}^{\text{inp}}(\omega) \quad (2.52)$$

which can be used to predict the time-dependent rate response.

Bibliography

- [1] Abbott LF & Kepler T (1990). Model Neurons: From Hodgkin-Huxley to Hopfield. In: *Statistical Mechanics of Neural Networks* (Garrido L, ed), pp 5-18. Berlin: Springer-Verlag.
- [2] Ahmed B, Anderson JC, Douglas RJ, Martin KA & Whitteridge D (1998). Estimates of the net excitatory currents evoked by visual stimulation of identified neurons in cat visual cortex. *Cereb Cortex*, **8**:462–476.
- [3] Amit DJ (1989). *Modeling Brain Function*. Cambridge: Cambridge UP.
- [4] Amit DJ & Tsodyks MV (1991). Quantitative Study of Attractor Neural Networks Retrieving at Low Spike Rates I: Substrate - Spikes, Rates and Neuronal gain. *Network*, **2**:259–273.
- [5] Azouz R, Gray CM, Nowak LG & McCormick DA (1997). Physiological properties of inhibitory interneurons in cat striate cortex. *Cereb Cortex*, **7**:534–545.
- [6] Ben-Yishai R, Lev Bar-Or R & Sompolinsky H (1995). Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci USA*, **92**:3844–3848.
- [7] Borg-Graham LJ, Monier C & Fregnac Y (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, **393**:369–373.
- [8] Brizzi L, Hansel D, Meunier C, van Vreeswijk C & Zytnicki D (2001). Shunting inhibition: a study using in vivo dynamic clamp on cat spinal motoneurons. *Soc Neurosci Abs*, 934.3.
- [9] Brunel N, Chance FS, Fourcaud N & Abbott LF (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys Rev Lett*, **86**:2186–2189.
- [10] Carandini M, Ferencsik M, Leonard CS & Movshon JA (1996). Spike train encoding by regular-spiking cells of the visual cortex. *J Neurophysiol*, **76**:3425–3441.

- [11] Carandini M, Anderson J & Ferster D (2000). Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J Neurophysiol*, **84**:909–926.
- [12] Chance FS, du Lac S & Abbott LF (2001). An integrate-or-fire model of spike rate dynamics. *Soc Neurosci Abs*, 821.44.
- [13] Chance FS, Abbott LF & Reyes AD (2002). Gain modulation from background synaptic input. *Neuron*, **35**:773–782.
- [14] Churchland PS & Sejnowski TJ (1992). *The Computational Brain*. Cambridge, MA: MIT Press.
- [15] Connors BW, Malenka RC & Silva LR (1988). Two inhibitory postsynaptic potentials, and GABAA and GABAB receptor-mediated responses in neocortex of rat and cat. *J Physiol (Lond)*, **406**:443–468.
- [16] Ermentrout B (1994). Reduction of conductance based models with slow synapses to neural nets. *Neural Comput*, **6**:679–695.
- [17] Ermentrout B (1998). Linearization of F-I curves by adaptation. *Neural Comput*, **10**:1721–1729.
- [18] Fourcaud N & Brunel B (2002). Dynamics of the Firing Probability of Noisy Integrate-and-Fire Neurons. *Neural Comput*, **14**:2057–2111.
- [19] Georgopoulos AP & Lukashin AV (1993). A dynamical neural network model for motor cortical activity during movement: population coding of movement trajectories. *Biol Cybern*, **69**:517–24.
- [20] Gerstner W (2000). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural Comput*, **12**:43–89.
- [21] Ginzburg I & Sompolinsky H (1994). Theory of correlations in stochastic neuronal networks. *Phys Rev E*, **50**:3171–3191.
- [22] Hansel D & Sompolinsky H (1996). Chaos and synchrony in a model of a hypercolumn in visual cortex. *J Comput Neurosci*, **3**:7–34.
- [23] Hansel D & Sompolinsky H (1998). Modeling feature selectivity in local cortical circuits. In: *Methods in Neuronal Modeling: from Synapses to Networks* (Koch C, Segev I, eds) 2nd edition, pp 499–567. Cambridge, MA:MIT Press.
- [24] Hille B (1984). Ionic channels of excitable membranes. Sunderland, MA: Sinauer.

- [25] Hodgkin AL & Huxley AF (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol (Lond)*, **117**:500–562.
- [26] Holt GR & Koch C (1997). Shunting inhibition does not have a divisive effect on firing rates. *Neural Comput*, **9**:1001–1013.
- [27] Hopfield JJ (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci USA*, **79**:1554–2558.
- [28] Kernell D (1968). The repetitive impulse discharge of a simple model compared to that of spinal motoneurons. *Brain Research*, **11**:685–687.
- [29] Knight BW (1972a). Dynamics of encoding in a population of neurons. *J Gen Physiol*, **59**:734–766.
- [30] Knight BW (1972b). The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. Experimental evidence for resonant enhancement in the population response. *J Gen Physiol*, **59**:767–778
- [31] Press WH, Flannery BP, Teukolsky SA & Vetterling WT (1988). *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge, UK: Cambridge UP.
- [32] Rinzel J & Frankel P (1992). Activity patterns of a slow synapse network predicted by explicitly averaging spike dynamics. *Neural Comput*, **4**:534–545.
- [33] Rolls ET & Treves A (1998). *Neural Networks and Brain Function*. Oxford University Press, Oxford.
- [34] Salinas E & Abbott LF (1996). A model of multiplicative neural responses in parietal cortex. *Proc Natl Acad Sci USA*, **93**:11956–11961.
- [35] Seung HS (1996). How the brain keeps the eyes still. *Proc Natl Acad Sci USA*, **93**:3339–13344.
- [36] Stafstrom CE, Schwindt PC & Crill WE (1984). Repetitive firing in layer V neurons from cat neocortex in vitro. *J Neurophysiol*, **52**:264–77.
- [37] Wilson HR & Cowan J (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J*, **12**:1–24.
- [38] Zhang K (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*, **16**:2112–2126.

Chapter 3

An Information Maximization Approach to Overcomplete and Recurrent Representations

3.1 Abstract

The principle of maximizing mutual information is applied to learning overcomplete and recurrent representations. The underlying model consists of a network of input units driving a larger number of output units with recurrent interactions. In the limit of zero noise, the network is deterministic and the mutual information can be related to the entropy of the output units. Maximizing this entropy with respect to both the feedforward connections as well as the recurrent interactions results in simple learning rules for both sets of parameters. The conventional independent components (ICA) learning algorithm can be recovered as a special case when there is an equal number of output units and no recurrent connections. The application of these new learning rules is illustrated on a simple two-dimensional input example.¹

¹This chapter was published as: Shriki O, Sompolinsky H, and Lee DD (2001). An Information Maximization Approach to Overcomplete and Recurrent Representations. In *Advances in Neural Information Processing Systems*, **13**, eds. Leen TK, Dietterich TG and Tresp V, (MIT Press, Cambridge, MA.), pp. 612-618.

3.2 Introduction

Many unsupervised learning algorithms such as principal component analysis, vector quantization, self-organizing feature maps, and others use the principle of minimizing reconstruction error to learn appropriate features from multivariate data [1, 2]. Independent components analysis (ICA) can similarly be understood as maximizing the likelihood of the data under a non-Gaussian generative model, and thus is related to minimizing a reconstruction cost [3, 4, 5]. On the other hand, the same ICA algorithm can also be derived without regard to a particular generative model by maximizing the mutual information between the data and a nonlinearly transformed version of the data [2]. This principle of information maximization has also been previously applied to explain optimal properties for single units, linear networks, and symplectic transformations [7, 8, 9].

In these proceedings, we show how the principle of maximizing mutual information can be generalized to overcomplete as well as recurrent representations. In the limit of zero noise, we derive gradient descent learning rules for both the feedforward and recurrent weights. Finally, we show the application of these learning rules to some simple illustrative examples.

3.3 Information Maximization

The “Infomax” formulation of ICA considers the problem of maximizing the mutual information between N -dimensional data observations $\{\mathbf{x}\}$ which are input to a network resulting in N -dimensional output signals $\{\mathbf{s}\}$ [2].

Here, we consider the general problem where the signals \mathbf{s} are M -dimensional with $M \geq N$. Thus, the representation is overcomplete because there are more signal components than data components. We also consider the situation where a signal component s_i can influence another component s_j through a recurrent interaction K_{ji} . As a network, this is diagrammed in Fig. 1 with the feedforward connections described by the $M \times N$ matrix W and the recurrent connections by the $M \times M$ matrix K . The network response \mathbf{s} is a deterministic function of the input \mathbf{x} :

$$s_i = g \left(\sum_{j=1}^N W_{ij} x_j + \sum_{k=1}^M K_{ik} s_k \right) \quad (3.1)$$

where g is some nonlinear squashing function. In this case, the mutual information between the

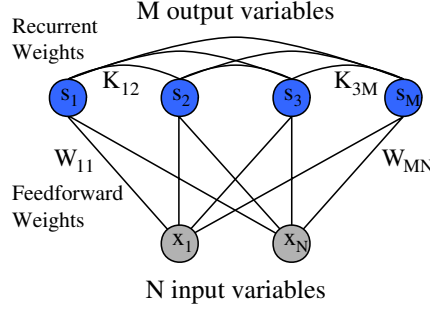


Figure 3.1: Network diagram of an overcomplete, recurrent representation. \mathbf{x} are input data which influence the output signals \mathbf{s} through feedforward connections W . The signals \mathbf{s} also interact with each other through the recurrent interactions K .

inputs \mathbf{x} and outputs \mathbf{s} is functionally only dependent on the entropy of the outputs:

$$I(\mathbf{s}, \mathbf{x}) = H(\mathbf{s}) - H(\mathbf{s}|\mathbf{x}) \sim H(\mathbf{s}). \quad (3.2)$$

The distribution of \mathbf{s} is a N -dimensional manifold embedded in a M -dimensional vector space and nominally has a negatively divergent entropy. However, as shown in Appendix 1, the probability density of \mathbf{s} can be related to the input distribution via the relation:

$$P(\mathbf{s}) \propto \frac{P(\mathbf{x})}{\sqrt{\det(\chi^T \chi)}} \quad (3.3)$$

where the susceptibility (or Jacobian) matrix χ is defined as:

$$\chi_{ij} = \frac{\partial s_i}{\partial x_j}. \quad (3.4)$$

This result can be understood in terms of the singular value decomposition (SVD) of the matrix χ . The transformation performed by χ can be decomposed into a series of three transformations: an orthogonal transformation that rotates the axes, a diagonal transformation that scales each axis, followed by another orthogonal transformation. A volume element in the input space is mapped onto a volume element in the output space, and its volume change is described by the diagonal scaling operation. This scale change is given by the product of the square roots of the eigenvalues of $\chi^T \chi$. Thus, the relationship between the probability distribution in the input and output spaces includes the proportionality factor, $\sqrt{\det(\chi^T \chi)}$, as formally derived in Appendix 1.

We now get the following expression for the entropy of the outputs:

$$H(\mathbf{s}) \sim - \int d\mathbf{x} P(\mathbf{x}) \log \left(\frac{P(\mathbf{x})}{\sqrt{\det(\chi^T \chi)}} \right) = \frac{1}{2} \langle \log \det(\chi^T \chi) \rangle + H(\mathbf{x}), \quad (3.5)$$

where the brackets indicate averaging over the input distribution.

3.4 Learning rules

From Eq. (3.5), we see that minimizing the following cost function:

$$E = -\frac{1}{2} \text{Tr} \langle \log(\chi^T \chi) \rangle, \quad (3.6)$$

is equivalent to maximizing the mutual information. We first note that the susceptibility χ satisfies the following recursion relation:

$$\chi_{ij} = g'_i \cdot \left(W_{ij} + \sum_k K_{ik} \chi_{kj} \right) = (GW + GK\chi)_{ij}, \quad (3.7)$$

where $G_{ij} = \delta_{ij} g'_i$ and $g'_i \equiv g' \left(\sum_j W_{ij} x_j + \sum_k K_{ik} s_k \right)$.

Solving for χ in Eq. (3.7) yields the result:

$$\chi = (G^{-1} - K)^{-1} W = \Phi W, \quad (3.8)$$

where $\Phi^{-1} \equiv G^{-1} - K$. Φ_{ij} can be interpreted as the sensitivity in the recurrent network of the i th unit's output to changes in the total input of the j th unit.

We next derive the learning rules for the network parameters using gradient descent, as shown in detail in Appendix 2. The resulting expression for the learning rule for the feedforward weights is:

$$\Delta W = -\eta \frac{\partial E}{\partial W} = \eta \langle \Gamma^T + \Phi^T \gamma \mathbf{x}^T \rangle \quad (3.9)$$

where η is the learning rate, the matrix Γ is defined as

$$\Gamma = (\chi^T \chi)^{-1} \chi^T \Phi \quad (3.10)$$

and the vector γ is given by

$$\gamma_i = (\chi \Gamma)_{ii} \frac{g''_i}{(g'_i)^3}. \quad (3.11)$$

Multiplying the gradient in Eq. (3.9) by the matrix (WW^T) yields an expression analogous to the “natural” gradient learning rule [10]:

$$\Delta W = \eta W \left(I + \langle \chi^T \gamma \mathbf{x}^T \rangle \right). \quad (3.12)$$

Similarly, the learning rule for the recurrent interactions is

$$\Delta K = -\eta \frac{\partial E}{\partial K} = \eta \langle (\chi \Gamma)^T + \Phi^T \gamma \mathbf{s}^T \rangle. \quad (3.13)$$

In the case when there are equal numbers of input and output units, $M = N$, and there are no recurrent interactions, $K = 0$, most of the previous expressions simplify. The susceptibility matrix χ is diagonal, $\Phi = G$, and $\Gamma = W^{-1}$. Substituting back into Eq. (3.9) for the learning rule for W results in the update rule:

$$\Delta W = \eta \left[(W^T)^{-1} + \langle \mathbf{z} \mathbf{x}^T \rangle \right], \quad (3.14)$$

where $z_i = g''_i / g'_i$. Thus, the well-known Infomax ICA learning rule is recovered as a special case of Eq. (3.9) [2].

3.5 Examples

We now apply the preceding learning algorithms to a simple two-dimensional ($N = 2$) input example. Each input point is generated by a linear combination of three (two-dimensional) unit vectors with angles of 0° , 120° and 240° . The coefficients are taken from a uniform distribution on the unit interval. The resulting distribution has the shape of a unit hexagon, which is slightly more dense close to the origin than at the boundaries. Samples of the input distribution are shown in Fig. 2. The second order cross correlations vanish, so that all the structure in the data is described only by higher order correlations. We fix the sigmoidal nonlinearity to be $g(x) = \tanh(x)$.

3.5.1 Feedforward weights

A set of $M = 3$ overcomplete filters for W are learned by applying the update rule in Eq. (3.9) to random normalized initial conditions while keeping the recurrent interactions fixed at $K = 0$. The length of the rows of W were constrained to be identical so that the filters are projections

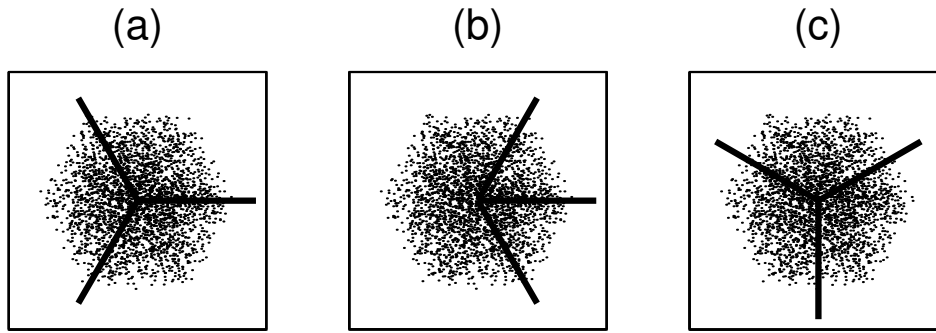


Figure 3.2: Results of fitting 3 filters to a 2-dimensional hexagon distribution with 10000 sample points.

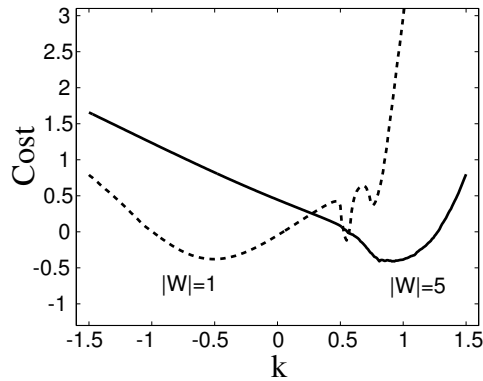


Figure 3.3: Effect of adding recurrent interactions to the representation. The cost function is plotted as a function of the recurrent interaction strength, for two different input scaling parameters.

along certain directions in the two-dimensional space. The algorithm converged after about 20 iterations. Examples of the resulting learned filters are shown by plotting the rows of W as vectors in Fig. 2. As shown in the figure, there are several different local minimum solutions. If the lengths of the rows of W are left unconstrained, slight deviations from these solutions occur, but relative orientation differences of 60° or 120° between the various filters are preserved.

3.5.2 Recurrent interactions

To investigate the effect of recurrent interactions on the representation, we fixed the feedforward weights in W to point in the directions shown in Fig. 2(a), and learned the optimal recurrent

interactions K using Eq. (3.13). Depending upon the length of the rows of W which scaled the input patterns, different optimal values are seen for the recurrent connections. This is shown in Fig. 3 by plotting the value of the cost function against the strength of the uniform recurrent interaction. For small scaled inputs, the optimal recurrent strength is negative which effectively amplifies the output signals since the 3 signals are negatively correlated. With large scaled inputs, the optimal recurrent strength is positive which tend to decrease the outputs. Thus, in this example, optimizing the recurrent connections performs *gain control* on the inputs.

3.6 Discussion

The learned feedforward weights are similar to the results of another ICA model that can learn overcomplete representations [11]. Our algorithm, however, does not need to perform approximate inference on a generative model. Instead, it directly maximizes the mutual information between the outputs and inputs of a nonlinear network. Our method also has the advantage of being able to learn recurrent connections that can enhance the representational power of the network. We also note that this approach can be easily generalized to undercomplete representations by simply changing the order of the matrix product in the cost function. However, more work still needs to be done in order to understand technical issues regarding speed of convergence and local minima in larger applications. Possible extensions of this work would be to optimize the nonlinearity that is used, or to adaptively change the number of output units to best match the input distribution.

We acknowledge the financial support of Bell Laboratories, Lucent Technologies, and the US-Israel Binational Science Foundation.

3.7 Appendix 1: Relationship between input and output distributions

In general, the relation between the input and output distributions is given by

$$P(\mathbf{s}) = \int d\mathbf{x} P(\mathbf{x}) P(\mathbf{s}|\mathbf{x}). \quad (3.15)$$

Since we use a deterministic mapping, the conditional distribution of the response given the input is given by $P(\mathbf{s}|\mathbf{x}) = \delta(\mathbf{s} - g(W\mathbf{x} + K\mathbf{s}))$. By adding independent Gaussian noise to the responses of the output units and considering the limit where the variance of the noise goes to zero, we can write this term as

$$P(\mathbf{s}|\mathbf{x}) = \lim_{\Delta \rightarrow 0} \frac{1}{(2\pi\Delta^2)^{N/2}} e^{-\frac{1}{2\Delta^2} \|\mathbf{s} - g(W\mathbf{x} + K\mathbf{s})\|^2}. \quad (3.16)$$

Any point in the output space can be decomposed into its projection on the image of the input space, \mathbf{s}^{\parallel} , and the orthogonal part, \mathbf{s}^{\perp} . Consider a point \mathbf{s} that lies inside the image ($\mathbf{s}^{\parallel} = \mathbf{s}$). This means that there exists \mathbf{x}_0 such that $\mathbf{s} = g(W\mathbf{x}_0 + K\mathbf{s})$. For small Δ , we can expand $g(W\mathbf{x} + K\mathbf{s}) - \mathbf{s} \simeq \chi\delta\mathbf{x}$, where χ is defined in Eq. (3.4), and $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$. We then get

$$\begin{aligned} P(\mathbf{s}|\mathbf{x}) &= \lim_{\Delta \rightarrow 0} \frac{1}{(2\pi\Delta^2)^{N/2}} e^{-\frac{1}{2}\delta\mathbf{x}^T \left(\frac{\chi^T \chi}{\Delta^2} \right) \delta\mathbf{x}} \\ &= \frac{1}{\sqrt{\det(\chi^T \chi)}} \left[\lim_{\Delta \rightarrow 0} \frac{e^{-\frac{1}{2}\delta\mathbf{x}^T \left(\frac{\chi^T \chi}{\Delta^2} \right) \delta\mathbf{x}}}{(2\pi)^{N/2} \sqrt{\det(\Delta^2 (\chi^T \chi)^{-1})}} \right]. \end{aligned} \quad (3.17)$$

The expression in the square brackets is a delta function in \mathbf{x} around \mathbf{x}_0 . Using Eq. (3.15) we finally get

$$P(\mathbf{s}) = \frac{P(\mathbf{x})}{\sqrt{\det(\chi^T \chi)}} \delta(\mathbf{s}^{\perp}). \quad (3.18)$$

where the factor $\delta(\mathbf{s}^{\perp})$ simply means that outside the image of the input space the density is zero. Note that for the case when χ is a square matrix ($M = N$), this expression reduces to the relation $P(\mathbf{s}) = P(\mathbf{x})/|\det(\chi)|$.

3.8 Appendix 2: Derivation of the learning rules

To derive the appropriate learning rules, we need to calculate the derivatives of E with respect to some set of parameters λ . In general, these derivatives are obtained from the expression:

$$\frac{\partial E}{\partial \lambda} = -\frac{1}{2} \text{Tr} \left\langle (\chi^T \chi)^{-1} \frac{\partial (\chi^T \chi)}{\partial \lambda} \right\rangle = -\text{Tr} \left\langle (\chi^T \chi)^{-1} \chi^T \frac{\partial \chi}{\partial \lambda} \right\rangle. \quad (3.19)$$

3.8.1 Feedforward weights

In order to derive the learning rule for the weights W , we first calculate

$$\frac{\partial \chi_{ab}}{\partial W_{lm}} = \sum_c \left(\Phi_{ac} \frac{\partial W_{cb}}{\partial W_{lm}} + \frac{\partial \Phi_{ac}}{\partial W_{lm}} W_{cb} \right) = \Phi_{al} \delta_{bm} + \sum_c \frac{\partial \Phi_{ac}}{\partial W_{lm}} W_{cb}. \quad (3.20)$$

From the definition of Φ , we see that:

$$\frac{\partial \Phi_{ac}}{\partial W_{lm}} = - \sum_{ij} \Phi_{ai} \frac{\partial G_{ij}^{-1}}{\partial W_{lm}} \Phi_{jc} \quad (3.21)$$

and

$$\frac{\partial G_{ij}^{-1}}{\partial W_{lm}} = - \frac{\delta_{ij}}{(g'_i)^2} \frac{\partial g'_i}{\partial W_{lm}} = - \delta_{ij} \frac{g''_i}{(g'_i)^3} \frac{\partial s_i}{\partial W_{lm}}, \quad (3.22)$$

where $g''_i \equiv g'' \left(\sum_j W_{ij} x_j + \sum_k K_{ik} s_k \right)$.

The derivatives of \mathbf{s} also satisfy a recursion relation similar to Eq. (3.7):

$$\frac{\partial s_i}{\partial W_{lm}} = g'_i \cdot \left(\delta_{il} x_m + \sum_j K_{ij} \frac{\partial s_j}{\partial W_{lm}} \right), \quad (3.23)$$

which has the solution:

$$\frac{\partial s_i}{\partial W_{lm}} = \Phi_{il} x_m. \quad (3.24)$$

Putting all these results together in Eq. (3.19) and taking the trace, we get the gradient descent rule in Eq. (3.9).

3.8.2 Recurrent interactions

To derive the learning rules for the recurrent weights K , we first calculate the derivatives of χ_{ab} with respect to K_{lm} :

$$\frac{\partial \chi_{ab}}{\partial K_{lm}} = \sum_c \frac{\partial \Phi_{ac}}{\partial K_{lm}} W_{cb} = - \sum_{c,i,j} \Phi_{ai} \frac{\partial \Phi_{ij}^{-1}}{\partial K_{lm}} \Phi_{jc} W_{cb}. \quad (3.25)$$

From the definition of Φ , we obtain:

$$\frac{\partial \Phi_{ij}^{-1}}{\partial K_{lm}} = - \frac{\delta_{ij}}{(g'_i)^2} \frac{\partial g'_i}{\partial K_{lm}} - \delta_{il} \delta_{jm}. \quad (3.26)$$

The derivatives of g' are obtained from the following relations:

$$\frac{\partial g'_i}{\partial K_{lm}} = \frac{g''_i}{g'_i} \frac{\partial s_i}{\partial K_{lm}} \quad (3.27)$$

and

$$\frac{\partial s_i}{\partial K_{lm}} = \Phi_{il} s_m. \quad (3.28)$$

which results from a recursion relation similar to Eq. (3.23). Finally, after combining these results and calculating the trace, we get the gradient descent learning rule in Eq. (3.13).

Bibliography

- [1] Jolliffe IT (1986). *Principal Component Analysis*. New York: Springer-Verlag.
- [2] Haykin S (1999). *Neural networks: a comprehensive foundation*. 2nd ed., Prentice-Hall, Upper Saddle River, NJ.
- [3] Jutte C & Herault J (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, **24**:1–10.
- [4] Hinton G & Ghahramani Z (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions Royal Society B*, **352**:1177–1190.
- [5] Pearlmutter B & Parr L (1996). A context-sensitive generalization of ICA. In ICONIP'96, 151–157.
- [6] Bell AJ & Sejnowski TJ (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Comput*, **7**:1129–1159.
- [7] Barlow HB (1989). Unsupervised learning. *Neural Comput*, **1**:295–311.
- [8] Linsker R (1992). Local synaptic learning rules suffice to maximize mutual information in a linear network. *Neural Comput*, **4**:691–702.
- [9] Parra L, Deco G & Miesbach S (1996). Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Comput*, **8**:260–269.
- [10] Amari S, Cichocki A & Yang H (1996). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, **8**:757–763.
- [11] Lewicki MS & Sejnowski TJ (2000). Learning overcomplete representations. *Neural Computation*, **12**:337–365.

Chapter 4

The Role of Recurrent Interactions in Orientation Selectivity: an Information Maximization Approach

4.1 Abstract

The computational role of recurrent connections in processing sensory inputs is investigated using a “ring model” of orientation selective neurons in visual cortex as an example. The optimal pattern of recurrent interactions is derived from an info-max principle using both analytical and numerical methods. When the characteristic contrast of the inputs to the network is small, the optimal profile of the recurrent interactions has a “Mexican hat” shape, with a modulation amplitude close to the symmetry breaking phase transition. This pattern of interactions greatly amplifies the response profile that is generated by the feedforward input. When the contrast is high the recurrent connections are inhibitory at the center and excitatory away from it. This pattern of connections reduces the saturation of the network and allows the neurons to use their full dynamic range. The results suggest that the functional nature of recurrent cortical connections may not be hard wired but dynamically adapted to the statistics of the sensory stimuli. ¹

¹This chapter was submitted to the *Neural Information Processing Systems* conference (NIPS 2003). Authors: Oren Shriki and Haim Sompolinsky.

4.2 Introduction

Recurrent connections are abundant in cortical circuitry but their functional role has been the subject of intense debates in cortical theory [1, 2]. In this paper we shed light on this problem by developing a first-principle information theoretic approach to investigate the computational role of recurrent connections in sensory processing. We focus on a hypercolumn in visual cortex consisting of neurons that process contrast and orientation information in a local patch of the visual field. Anatomical and physiological studies indicate that a substantial part of the synaptic input to each neuron in the hypercolumn comes from other neurons in the hypercolumn. However, the functional role of these recurrent interactions is still unresolved. Of particular interest are the questions (1) are the cortical recurrent connections predominantly suppressive or amplifying and (2) in what way do the recurrent interactions depend on the preferred orientations (POs) of the connected pairs. Here we consider a simplified model of a hypercolumn, with a one-dimensional architecture - a “ring network”, and use the principle of maximum mutual information to evaluate the optimal pattern of recurrent interactions in such a network and its dependence on the statistics of the external inputs. To derive the optimal pattern of connections we apply the general info-max approach for overcomplete and recurrent representations developed in [3].

4.3 Network Structure

Each input sample is a point on the plane, with an angle, θ_0 , representing the orientation of a visual stimulus and amplitude (its distance from the origin), r , representing the stimulus contrast $(x_1, x_2) = r(\cos\theta_0, \sin\theta_0)$. For clarity and generality of presentation we consider a periodicity of 360° (rather than 180° , which is the relevant symmetry when considering orientations). The angles θ_0 are distributed uniformly between 0 and 2π . The amplitudes r are distributed according to a Gaussian distribution with a positive mean $\langle r \rangle$. By varying the mean value of r we will study the effect the stimulus statistics on the optimal network connections (see below). The network represents this two-dimensional input by N sigmoidal neurons ($N \gg 1$) interconnected with recurrent interactions (K), as shown in Fig. 4.1. The feedforward connections (the rows of W) are unit vectors, uniformly distributed over all possible directions, i.e $(W_{i1}, W_{i2}) = (\cos\phi_i, \sin\phi_i)$ where

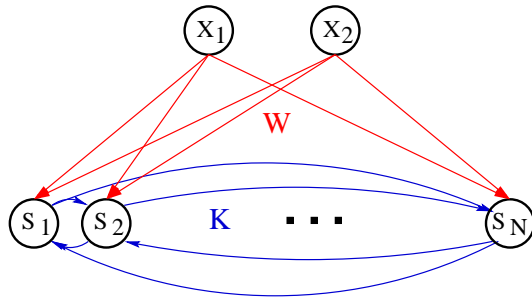


Figure 4.1: Network architecture. The network contains two input neurons and N output neurons. The feedforward connections are unit vectors, uniformly distributed over all possible directions. The output neurons are connected by recurrent interactions.

$\phi_i = 2\pi i/N$, $i = 1, \dots, N$. Thus, the input to each neuron is tuned to the angle of the stimulus with a cosine tuning function, which is peaked at ϕ_i , the PO of the i 'th neuron. This type of feedforward connections represents the orientation tuning of the input from the Lateral Geniculate Nucleus to neurons in the primary visual cortex. Our goal is to find the optimal matrix of recurrent connections K .

Our optimality criterion for K will be to maximize the mutual information between the *steady state* responses of the neurons s_i and the stimulus. For a given input and connections K , the steady-state responses are given by rate model fixed point equations,

$$s_i = g \left(\sum_j K_{ij} s_j + x_1 \cos \theta_i + x_2 \sin \theta_i \right) \quad (4.1)$$

where g is a smooth monotonic non-linear transfer function. A concrete example is $\tanh(x)$. Note that with this choice the neuronal activity ranges from -1 to $+1$. Also, when the total input is zero the activity is zero as well, namely the neuron is in the middle of its dynamic range. This is analogous to the assumption that the neurons have considerable spontaneous activity.

4.4 The Cost Function

Equation (4.1) describes a deterministic relation between \mathbf{s} and the input \mathbf{x} . In these low noise conditions, the mutual information between the inputs and outputs depends only on the entropy of the outputs $H(\mathbf{s})$. As shown in [3], maximizing this entropy is equivalent to minimizing the

following cost function:

$$E = -\frac{1}{2} \text{Tr} \langle \log(\chi^T \chi) \rangle = -\frac{1}{2} \langle \log \det(\chi^T \chi) \rangle \quad (4.2)$$

where the brackets indicate averaging over the input distribution. The sensitivity matrix (also termed the Jacobian or susceptibility matrix), χ , is an $N \times 2$ matrix given by:

$$\chi_{i1} = \frac{\partial s_i}{\partial x_1} = g'_i \cdot \left[\sum_l K_{il} \chi_{l1} + \cos(\theta_i) \right] \quad (4.3)$$

$$\chi_{i2} = \frac{\partial s_i}{\partial x_2} = g'_i \cdot \left[\sum_l K_{il} \chi_{l2} + \sin(\theta_i) \right] \quad (4.4)$$

where $g'_i \equiv g'(\sum_j K_{ij} s_j + x_1 \cos \theta_i + x_2 \sin \theta_i)$ is the derivative function of the neuronal transfer function and we have used the expression for s_i given in Eq. (4.1).

In [3] a gradient-based algorithm has been derived for the numerical minimization of E with respect to K . This algorithm can be applied to the present problem for a general stimulus ensemble. Here we focus on two interesting limits. In the low contrast case, the feedforward component of the input to each neuron is small. Under these conditions, without recurrent connections the neurons utilize only a small part of their dynamic range near 0. The other extreme is when the typical input contrast is very high, in which case the feedforward component drives the neuronal activities into their saturating limits (either -1 or $+1$ in our model). Under these conditions, the neuronal response without recurrent connections cannot reflect small changes in the input. We will study below how the recurrent connections improve the neuronal representation of the input in these two limits.

4.5 The Low Contrast Case

We first investigate the optimal pattern of recurrent interactions when the typical input contrast is low, namely $r \rightarrow 0$.

4.5.1 Analytical derivation of the optimal pattern of interactions

In the analytical derivation we assume that the interaction K_{ij} between the i 'th and j 'th neurons is an even function of the distance between the neurons on the ring,

$$K_{ij} = K(\theta_i - \theta_j) \quad (4.5)$$

When r approaches 0 the total external input to each neuron approaches zero. We denote the value of g' at zero input by $\gamma_0 \equiv g'(0)$. In the case of \tanh , $\gamma_0 = 1$. Since the number of output neurons, N , is large we can take the continuum limit and transform the summations over the angles to integrals. For instance, the equation for χ_{i1} can be written as

$$\chi_1(\theta) = \gamma_0 \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} d\theta' K(\theta - \theta') \chi_1(\theta') + \cos \theta \right] \quad (4.6)$$

and similarly for χ_{i2} . We define the Fourier series of K and χ_1

$$K(\theta) = \frac{1}{N} \sum_{n=0}^{\infty} k_n \cos(n\theta) \quad (4.7)$$

$$\chi_1(\theta) = \sum_{n=0}^{\infty} [a_n \cos(n\theta) + b_n \sin(n\theta)]. \quad (4.8)$$

Fourier transforming Eq. (4.6) yields $a_n = \gamma_0 \delta_{n1} / (1 - \frac{1}{2} \gamma_0 k_1)$ and $b_n = 0$, where k_1 is the first cosine harmonic of the interaction profile, Eq. (4.7). Thus,

$$\chi_{i1} = \frac{\gamma_0 \cos \theta_i}{1 - \frac{1}{2} \gamma_0 k_1} \quad (4.9)$$

and similarly

$$\chi_{i2} = \frac{\gamma_0 \sin \theta_i}{1 - \frac{1}{2} \gamma_0 k_1}. \quad (4.10)$$

The 2×2 matrix $\chi^T \chi$ is a diagonal matrix with elements

$$\left(\chi^T \chi \right)_{11} = \left(\chi^T \chi \right)_{22} = \frac{1}{2 \left(1 - \frac{1}{2} \gamma_0 k_1 \right)^2}. \quad (4.11)$$

Substituting these expressions in Eq. (4.2), yields

$$E = \frac{1}{2} \log 4 + 2 \log \left(1 - \frac{1}{2} \gamma_0 k_1 \right) \quad (4.12)$$

This cost function is shown in Fig. 4.2 together with a numerical evaluation of the cost function using the full network dynamics.

Equation (4.12) implies that as k_1 approaches the critical value $k_1^c = 2/\gamma_0$ the cost function diverges to $-\infty$. This means that the optimal recurrent interactions are of the form

$$K(\theta_i - \theta_j) = \frac{2}{N\gamma_0} \cos(\theta_i - \theta_j) \quad (4.13)$$

The divergence of the sensitivity (or susceptibility) at k_1^c reflects the fact that at this point the network undergoes a phase-transition into a state of *spontaneous symmetry breaking* [5]. For $k_1 < 2/\gamma_0$, if the input has zero contrast ($r = 0$), the network settles into a homogeneous state with $s_i = g(0)$. However, for $k_1 > 2/\gamma_0$, the network dynamics evolve into an inhomogeneous solution with a typical “hill” shape, which is determined by the recurrent connections. The location of the peak of this hill is arbitrary and depends on the specific realization of the noise in the input pattern and on the initial conditions of the neuronal activities. This dramatic change in the network behavior implies that near k_1^c the network is extremely sensitive to small changes in the input. This enhanced sensitivity increases the mutual information between the network responses and the stimulus.

Amplification of external inputs

In the limit of low contrasts the amount of amplification of the input by the optimal recurrent network can be estimated analytically. Placing the optimal interaction profile in Eq. (4.1) and taking the continuum limit we obtain (here we use for simplicity $g(x) = \tanh(x)$),

$$\begin{aligned} s(\theta) &= \tanh\left(r \cos(\theta) + 2 \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} \cos(\theta - \theta') s(\theta')\right) \\ &= \tanh[(r + 2m_1) \cos(\theta)] \end{aligned} \quad (4.14)$$

where $m_1 \equiv \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} \cos(\theta') s(\theta')$.

We have assumed, without loss of generality, that the angle of the external stimulus is $\theta_0 = 0$. Expanding the tanh in a Taylor series to third order yields

$$s(\theta) = (r + 2m_1) \cos(\theta) - \frac{1}{3}(r + 2m_1)^3 \cos^3(\theta). \quad (4.15)$$

Multiplying by $\cos(\theta)$ both sides and integrating over θ we obtain

$$m_1 = \frac{r}{2} + m_1 - \frac{1}{8}(r + 2m_1)^3. \quad (4.16)$$

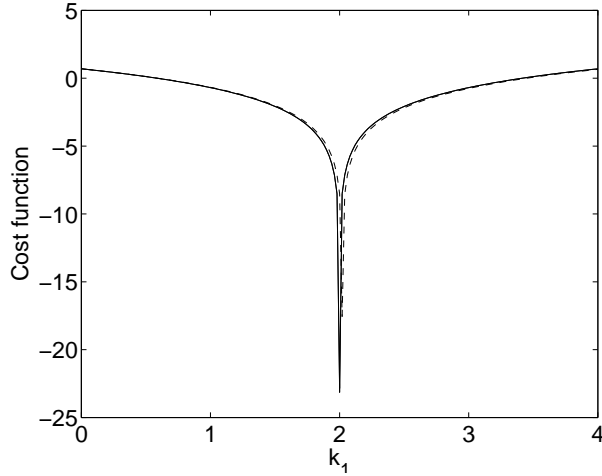


Figure 4.2: The cost function as a function of the first harmonic of the interaction profile. The solid line represents the result from the analytical expression Eq. (4.12). The dashed line results from a direct calculation of the full cost function, Eq. (4.2). For the full cost function we simulated the dynamics of a network with $N = 200$ neurons. The feedforward connections were unit vectors in the plane uniformly distributed between 0° and 360° . The recurrent interactions had the form $K_{ij} = \frac{k_1}{N} \cos(\theta_i - \theta_j)$ and for each value of k_1 we ran the network dynamics over a set of 2000 random inputs. The angles of the inputs were uniformly distributed on the circle and the magnitude, r , was drawn from a Gaussian distribution with mean 0.001 and standard deviation 0.0001.

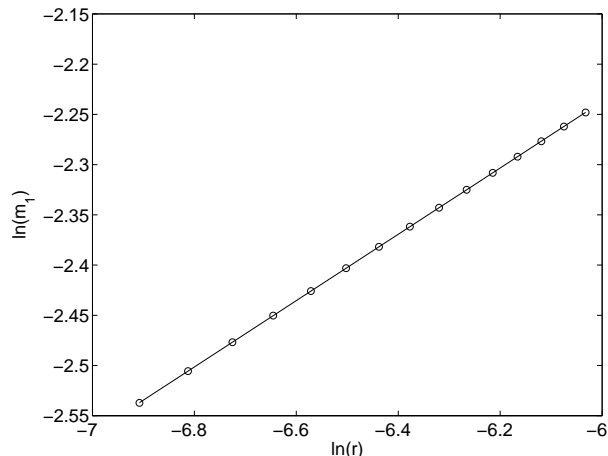


Figure 4.3: The \log of m_1 as a function of the \log of r for very low values of r . See text for details.

Extracting m_1 and using the fact that r is small compared to m_1 yields

$$m_1 = \left(\frac{r}{2}\right)^{\frac{1}{3}}. \quad (4.17)$$

This should be compared with a feedforward network, in which $m_1 \sim r$. Thus, the amplification by the recurrent interactions is reflected in the generation of a nonlinear response to low contrast inputs.

This result is verified in simulations of the network dynamics, see Fig. 4.3. The network contained 141 neurons with recurrent interactions $K_{ij} = \frac{2}{N} \cos(\theta_i - \theta_j)$. For each value of r we simulated the dynamics of the network and calculated m_1 (r varied between 0.001 and 0.0024). The circles in Fig. 4.3 show the \log of m_1 against the \log of r . The solid line is the optimal linear fit, and its slope is 0.3304, in agreement with Eq. (4.17).

4.5.2 Numerical simulations of the learning process

In the limit of $r \rightarrow 0$ the cost function depends solely on the first harmonics of K , leaving open the question of whether the higher order corrections in r predict large values of the higher harmonics of K . Furthermore, in the analytic derivation we have assumed translational invariance of K , which raises the question of whether there are better solutions which break this symmetry of K . To address these questions we simulated the gradient based learning algorithm for the evolution of

the interaction matrix that was developed in [3], with no restrictions on its form. The network consisted of 2 input neurons and 141 output neurons. The feedforward connections to each output neuron were unit vectors uniformly distributed between 0° and 360° , and were fixed throughout the learning. The initial recurrent interaction matrix was set to zero. The angle of each input was drawn from a uniform distribution, while the magnitude was drawn from a Gaussian distribution around a characteristic radius of 0.01 with a standard deviation of 0.001. The simulations were performed with a fixed learning rate, η .

Figure 4.4a shows the resulting interaction matrix. As can be seen, the interaction pattern is translation invariant; i.e., each neuron has the same pattern of interactions. It is important to note that we do not impose any symmetry on the connections. The resulting translation invariance is a natural result of the statistical symmetry of the inputs to the network. Figure 4.4b shows one column of the interaction matrix (the postsynaptic connections of a single neuron). The values are multiplied by the number of neurons, N , for clarity. In the Fourier analysis of the interaction profile the dominant contribution is of $\cos(\theta)$, with an amplitude $k_1 = 2.0164$ and the second dominant is of $\cos(3\theta)$, with an amplitude $k_3 = 0.0016$. This result is highly congruent with the analytical derivation presented above, Eq. (4.13), that predicts a pure cosine profile with an amplitude of 2. Figure 4.4c shows the response of the network (solid line) to a low contrast input ($r = 0.01$). The amplification in comparison to the network response without the recurrent interactions (dashed line) is clearly seen. The numerical results indicate that the higher harmonics of K vanish as $r \rightarrow 0$.

4.6 The High Contrast Case

We next study the pattern of recurrent interactions in the high contrast case. When the input distribution is concentrated around a high mean value, the feedforward input tends to saturate the neuronal activities. The learned recurrent interactions in this case are expected to “pull out” the neurons from saturation. Analytical study of this highly nonlinear regime is difficult. Here we report the results of numerical simulations.

Figure 4.5a shows the interaction matrix when the input magnitude r is drawn from a Gaussian distribution around 3 with a standard deviation of 0.25. The angles were drawn from a uniform

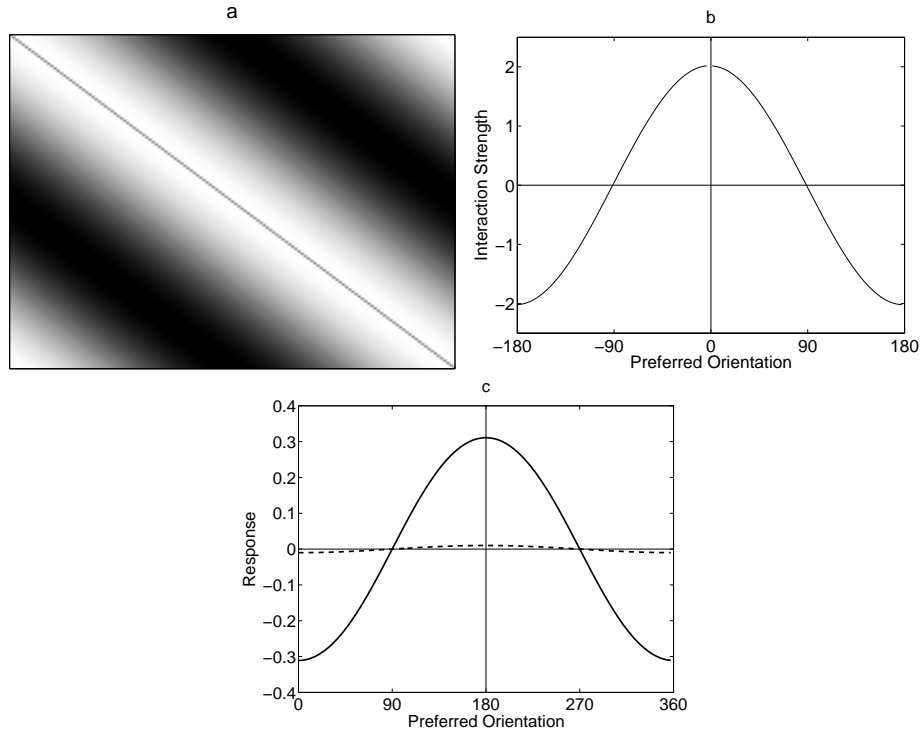


Figure 4.4: Results from numerical simulations of a gradient-based learning algorithm that minimizes the cost function, Eq.(4). The details of the algorithm are described in [3]. Inputs have mean contrast $r = 0.01$. a: Interaction matrix. The gray level at the i 'th row and j 'th column represents the value of the interaction from neuron j onto neuron i . The gray levels are scaled such that white represents the maximal value while black represents the minimal value. b: Interaction profile for the neuron that is tuned to 180° (the middle column of the interaction matrix). c: Network response for an input with $r = 0.01$. The dashed line is the response of the network without the recurrent interactions and the solid line is the response with them.

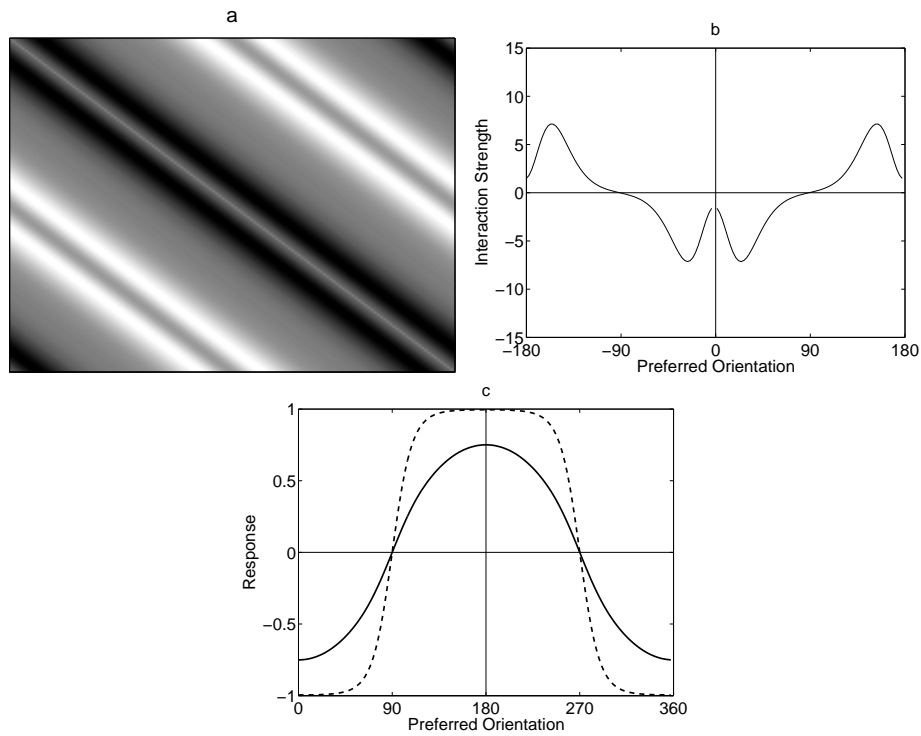


Figure 4.5: Results of numerical simulations of the learning process with mean input contrast $r = 3$. Here and in the following figure we have used time-dependent learning rate: $\eta(n) = 1/(\lambda_1 + \lambda_2(n - 1))$. a: Interaction matrix. b: Interaction profile for the neuron that is tuned to 180° . c: Network response for an input with $r = 3$. The dashed line is the response of the network without the recurrent interactions and the solid line is the response with them.

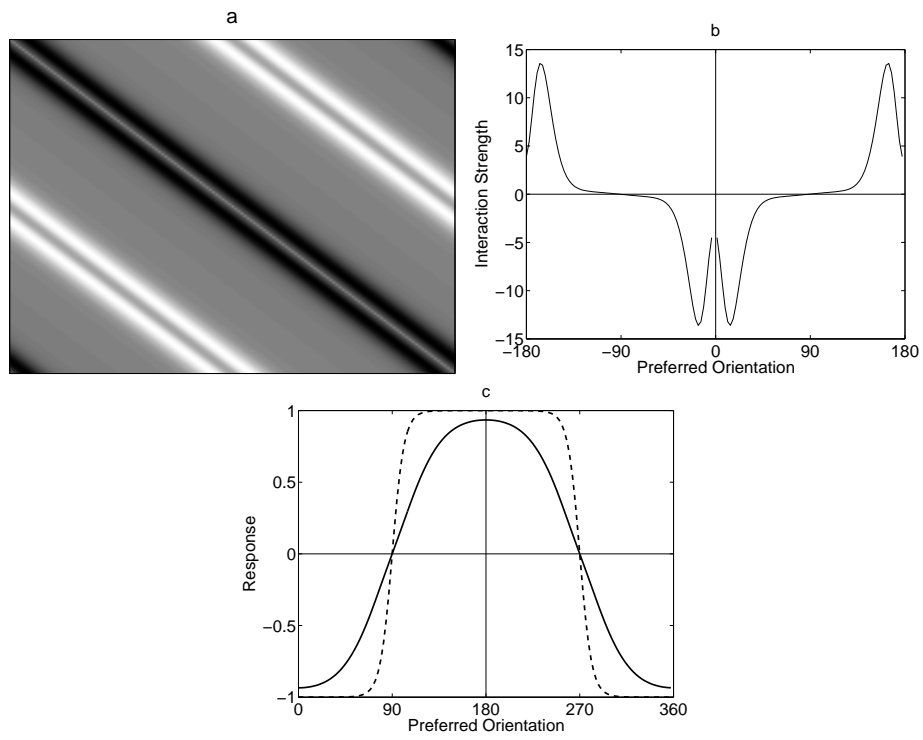


Figure 4.6: Results of numerical simulations with mean input contrast $r = 5$. (See caption of Fig. 4.5).

distribution and the number of neurons was 141. As in the low contrast case, the resultant pattern of interaction is translationally invariant reflecting the symmetry of the input statistics. Fig. 4.5b shows the learned interaction profile. It has a short range negative component and a long range positive one. This pattern of interactions acts to suppress the activities of neurons that are close to their upper saturation (+1) and to enhance the activities of neurons that are close to their lower saturation (-1). The net effect of this push-pull mechanism is clearly seen in Fig. 4.5c that compares the activity profile of the network with and without the recurrent interactions. The activity profile of the network is very close to a cosine profile. In the Fourier analysis of the network response, $s(\theta)$, the coefficient of $\cos(\theta)$ is the dominant one with a value of 0.7938, while the second dominant component is the coefficient of $\cos(3\theta)$ with a value of -0.0523. Note that the interaction profile is not simply an inverted Mexican hat but has an additional structure, namely a secondary peak at zero, see Fig. 4.5b. This structure is connected to a mirror symmetry of K about 90 degrees.

Similar results are obtained with other values of the mean radius that tend to drive the network into saturation. For instance, Fig. 4.6 shows the resulting interactions and the network response when $r = 5$. As in the case of $r = 3$ the activity profile of the network is very close to a cosine profile. In the Fourier analysis the coefficient of $\cos(\theta)$ is the dominant one with a value of 1.009, and the second dominant is the coefficient of $\cos(3\theta)$ with a value of -0.0726.

4.7 Discussion

We studied the functional role of recurrent connections in a simplified model of a local network in the primary visual cortex. We fixed the feedforward connections and evaluated the pattern of recurrent connections that optimizes the mutual information between the input and output of the network in the limit of low noise. This was done using a previously derived general method of evaluating optimal patterns of connections in networks with overcomplete and recurrent architectures [3]. This method generalized the ICA approach which is usually restricted to complete representations. ICA methods have been used to gain insight into the nature of the (essentially) feedforward receptive fields in visual cortex [4]. The present work can be viewed as a step towards extending this approach to gain insight into the recurrent connectivity in V1.

We have found that when the inputs are well inside the dynamical range of the neuronal re-

sponse, i.e, they do not lead to saturation, the network uses the recurrent interactions to amplify the inputs. Our results can be compared with the recurrent models of Ben-Yishai et al. [23] and others on a hypercolumn in V1 [1, 2, 6]. In some of these models, the main assumption is that the “net” interactions between orientation columns have a “Mexican hat” shape. Our model does not assume it but rather predicts that this should be the case from a first-principle information theoretic approach, for inputs which are below saturation. In addition, in order to explain the narrow tuning width and the contrast invariance of orientation tuning, Ben-Yishai et al. suggest that the network operates in the symmetry breaking regime, namely $k_1 \geq 2$. Here, we show that an optimal network will be close to the symmetry breaking transition point, again for low contrast inputs.

For high contrast inputs, we find that the optimal pattern of interactions acts to pull the neurons from saturation. The network response has a smooth peak in contrast to the square-wave feedforward response. This is achieved by a pattern of connections that is negative for short distances and positive for large distances. These connections tend to balance the excitatory and inhibitory components of the input to the neurons so that the net inputs span the linear part of their dynamic range.

Bibliography

- [1] Sompolinsky H & Shapley R (1997). New perspectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology*, **7**:514–522.
- [2] Ferster D & Miller KD (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Reviews of Neuroscience*, **23**:441–471.
- [3] Shriki O, Sompolinsky H, & Lee DD (2001). An Information Maximization Approach to Overcomplete and Recurrent Representations. *Advances in Neural Information Processing Systems*, **13**:612–618.
- [4] Bell AJ & Sejnowski TJ (1997). The ‘independent components’ of natural scenes are edge filters. *Vision Research*, **37**:3327–3338.
- [5] Ben-Yishai R, Lev Bar-Or R & Sompolinsky H (1995). Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci USA*, **92**:3844–3848.
- [6] Kang K, Shelley M & Sompolinsky H (2002). Mexican hats and pinwheels in visual cortex. *Proc Natl Acad Sci USA*, **100**:2836–2841.

Chapter 5

Discussion and conclusions

We have studied two central issues in modeling of recurrent neuronal networks. One concerns providing a description for the firing rates in a network of conductance-based neurons. The other concerns the role of recurrent interactions in the representation of sensory inputs. This chapter summarizes the main points from the discussions of the previous chapters and presents an overview of the entire work.

5.1 Rate Models for Conductance-based Networks

We have shown how to construct rate models that are appropriate for large networks of conductance-based neurons in which the network state does not possess a high degree of synchrony. Basically, we use an averaging method to show that in such a network the input to a cell can be described in terms of the presynaptic firing rates. We then use a simple phenomenological model, which is an extension of the concept of frequency-current (f-I) curves, to calculate the output firing rates. This results in a description of the network dynamics in terms of the neuronal firing rates rather than the voltages.

The most important assumptions are: 1) the independence of the gain of the f-I curve of the leak conductance, and 2) the approximated linear dependence of the threshold current on the leak conductance. A further simplifying assumption is that in a broad range of input currents and output firing rates the f-I curves of cortical neurons can be well approximated by a threshold linear

function. These assumptions are supported by experimental data.

The mapping of the conductance-based model onto a rate model described in this work provides a correspondence between the “synaptic efficacy” used in rate models and biophysical parameters of the neurons. In particular, the sign of the synaptic efficacy is determined by the value of the reversal potential relative to $E_L + V_c$, where E_L is the reversal potential of the leak current and V_c is the threshold gain-potential of the post-synaptic neuron (see Chapter 2). In rate models, a positive interaction characterizes a presynaptic source that tends to increase the firing rate of the postsynaptic neuron. In general, this does not coincide with the standard biological definition, in which the synaptic reversal potential is compared with the voltage threshold of the cell. It may be the case that inputs from a presynaptic neuron tend to elicit a single spike, but the accompanying increase in conductance will result in a reduction of the overall firing rate. Our theory allows for using rate models to calculate synaptic conductances that are generated by the recurrent activity. This enables quantitative comparison of predictions from rate models and conductance measurements in *in-vitro* and *in-vivo* intra-cellular experiments.

The model is demonstrated on several examples. In the case of a fully connected network of excitatory neurons we have shown that the firing rate of the neurons predicted by the rate model was highly congruent with simulation results in a broad range of synaptic conductances. This indicates that our rate model provides reliable results over a broad range of firing rates and conductance changes. We have also shown that our approach can be applied to networks with more complicated architectures, such as the conductance-based model of a hypercolumn in the primary visual cortex. An important result is that the rate model can be used to predict the stability conditions of the network. When the input to the network is uniform, the rate model accurately predicts what is the critical value of the modulation of the recurrent interactions for which the network experiences a symmetry-breaking phase-transition. Below this value, the network settles into a homogeneous state, while above it, the network converges to a profile of activity peaked at some orientation. In addition, when the stimulus has a characteristic orientation, the model accurately predicts the firing rate profile of the neurons as well as the magnitude of the conductance change in each neuron induced by the stimulus. Our theory is quite general and can be applied to other brain areas besides the visual cortex. Indeed, a recent theoretical study of oculomotor neuronal integration has utilized

key elements of our model [1].

To extend our approach to the case of time dependent inputs, we have studied the response of the single neuron to an input which is modulated sinusoidally in time. We have shown that this response can be described by a rate model in which the neuron responds instantaneously to an effective input, which is a filtered version of the instantaneous one. This is similar to the approach of Chance et al (2001), who studied the responses of spiking neurons to oscillating input currents. However, in contrast to Chance et al. we find that the time filter more complicated than a simple low-pass or high-pass filter. Our description is valid provided that the input is sufficiently noisy to broaden resonances and that its modulation is small compared to its average to avoid rate rectification. Interestingly, we have found that if these assumptions are satisfied the neuron essentially behaves like a linear device even if the modulation of the input is substantial. This allowed us to derive a rate model, which provides a good description for the dynamics of a homogeneous network of neurons receiving a time dependent input. However, applications of our approach to other single-cell conductance-based models and to more complicated network architectures still need to be explored.

Lastly, we have focused in this work on conductance-based network models of point neurons. An interesting open question is whether an appropriate rate model can also be derived for neurons with extended morphology. In conclusion, our results open up possibilities for applications of rate models to a range of problems in sensory and motor neuronal circuits in cortex as well as in other structures with similar single neuron properties. Applying the rate model to a concrete neuronal system requires knowledge of a relatively small number of experimentally measurable parameters that characterize the f-I curves of the neurons in the system. In addition, the rate model requires knowledge of the gross features of the underlying connectivity pattern, as well as an estimate of the order of magnitude of the associated synaptic conductances.

5.2 Optimizing Recurrent Network Architectures

In Chapter 3 we apply the principle of maximizing mutual information to learning of overcomplete and recurrent neuronal representations of sensory stimuli. The work is novel in two ways. The first is derivation of the cost function for a feedforward network with an overcomplete architecture. This is an important generalization of the standard ICA (Independent Component Analysis) approach,

which is usually limited to complete representations [2]. In fact, the framework that we propose applies to undercomplete representations as well. This is a significant contribution, since complete representations, in which the number of output units equals the number of input units, are very limited and in most cases seem artificial.

The second novel feature is the derivation of a learning rule for the recurrent connections. This involves a solution of self-consistent equations and, as we show, results in a compact learning rule. We demonstrate through a simple example that recurrent interactions can enhance the representational power of the network by performing gain control on the inputs. In other words, they prevent the neurons from reaching saturation.

The work can also be used in technological applications. A common application of the standard ICA algorithm is to the problem of blind source separation. In this problem an array of receivers picks up linear mixtures of a number of source signals. The task is to estimate the original source signals from these mixtures. The standard ICA approach is limited since the number of receivers has to be the same as the number of sources. Our approach can be applied even when there are fewer mixtures than sources. Moreover, the recurrent interactions can be used to characterize the interactions between the sources. This relieves the restriction of independent sources.

Importantly, the mathematical procedures that are presented in this work are very general and can also be applied to derive learning algorithms from other computational approaches. For instance, the cost function can be formulated using the concept of *Fisher information* rather than mutual information, yet the derivation of the learning rules can be done along the same mathematical lines.

From a broader perspective, the work can be seen as completing the scheme that Linsker proposed in 1988 [3]. Linsker introduced the principle of maximizing the mutual information between the input and output of a neuronal network, so called the *infomax* principle, and studied feedforward linear networks under certain assumptions about the input statistics, such as Gaussianity. The standard ICA algorithm generalized the idea to networks of nonlinear units with arbitrarily distributed inputs, but with the constraint of the same number of input and output units. Our approach provides a further generalization to networks of nonlinear units with the number of output units larger or smaller than the number of input units and with recurrent interactions in the

output layer.

However, more work still needs to be done in order to understand technical issues regarding speed of convergence and local minima in the learning algorithm in large scale applications. Possible extensions of this work would be to optimize the nonlinearity that is used, or to adaptively change the number of output units to best match the input distribution. Another extension of the work would be to describe the neuronal dynamics using more realistic rate models. For instance, it is possible to incorporate the second-order linear filter that we propose in Chapter 2.

5.3 The Role of Recurrent Interactions in Orientation Selectivity: an Information Maximization Approach

As an application of the information maximization approach we studied the functional role of recurrent interactions in a network model of a hypercolumn in the primary visual cortex. The input to the network was characterized by its angle and its contrast and was then represented by a large number of output neurons. The feedforward connections were set so that each neuron had a tuning to a different angle. Thus the output layer could be viewed as a simplified model of a hypercolumn. We then studied the pattern of recurrent interactions that optimizes the mutual information between the input and output of the network.

In the limit of low contrasts, the optimal profile of interactions between orientation columns has a "Mexican hat" shape, which helps to amplify the inputs to the network and span more efficiently the neuronal dynamic range. The maximal amplification is achieved when the modulation amplitude of this profile approaches a critical value, for which the network undergoes a phase transition. Below this value the response to a homogeneous input is also homogeneous, while above it the response to the same input becomes inhomogeneous, peaking at a certain orientation. The behavior above the transition is termed *spontaneous symmetry breaking*.

Natural scenes are characterized by an exponential probability of contrasts, which means that under normal conditions the low contrasts are dominant. Thus, our model supports the notion that intracortical interactions have a dominant role in shaping the orientation tuning properties of neurons in the primary visual cortex. Moreover, it predicts that the interaction profile will have a

“Mexican hat” shape with a characteristic modulation amplitude such that the network operates near the symmetry breaking regime.

We have also shown that for relatively high contrast inputs the recurrent interactions develop a different pattern, intended to prevent the network from reaching saturation. These connections tend to balance the excitatory and inhibitory components of the input to the neurons so that the net inputs span the linear part of their dynamic range. This idea has some practical implementations. Consider an array of detectors with a limited sensitivity range. As shown, recurrent interactions among the different detectors could enhance the sensitivity of the array. Of course, for the learning algorithm to be useful the input patterns should have some characteristic statistical properties, which is often the case.

It is interesting that in the present model the recurrent connections preserve the underlying isotropy of the stimulus ensemble. It would be interesting to explore conditions under which this symmetry is spontaneously broken by the learning process leading e.g., to directional patterns of recurrent connections.

5.4 A First-principle Computational Approach to Quantitative Biophysical Predictions

The two main themes of this thesis can be combined to form a coherent approach to the modeling of sensory neuronal systems. The first stage would be to construct a rate model description for the underlying system. This network model could contain both feedforward and recurrent interactions. The learning algorithms that we present in Chapter 3 could then be applied to find the optimal pattern of interactions. This should be done using inputs that are statistically similar to the inputs to the original system. The second stage would be to use the mapping between rate models and conductance-based models developed in Chapter 2 to translate the results into quantitative biophysical predictions about the system.

For instance, the prediction for the optimal pattern of recurrent interactions in the hypercolumn model of Chapter 4 can be translated into predictions about the firing rates and conductance changes in the cortical network. In fact, such predictions appear in Chapter 2.

A more complicated application would be to the issue of long-range interactions in the primary visual cortex. The term *long range interactions* refers in this context to interactions between neurons with non-overlapping receptive fields. The inputs should be natural images (or even more generally natural movies) while the output layer represents the neuronal activities in a set of hypercolumns. After the learning stage, the responses of the network can be compared to responses of neurons in V1 to similar inputs. It is well known that the response of neurons in V1 to inputs that extend beyond their classical receptive field is strongly nonlinear, and seem to involve recurrent processing inside V1 [4, 5]. Thus, it would be interesting to compare these responses to the responses of the optimal recurrent network model derived by our method. Moreover, using the mapping prescription, the model can predict the conductance change in a neuron induced by general stimuli that cover more than the cell's receptive field.

Bibliography

- [1] Seung SH, Lee DD, Reis BY & Tank DW (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron*, **26**:259-271.
- [2] Bell AJ & Sejnowski TJ (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Comput*, **7**:1129–1159.
- [3] Linsker R (1988). Self-organization in a perceptual network. *IEEE Computer*, **21**:105–117.
- [4] Gilbert CD (1992). Horizontal integration and cortical dynamics. *Neuron*, **9**: 1–13.
- [5] Toth LJ, Rao SC, Kim D, Somers D & Sur M (1996). Subthreshold facilitation and suppression in primary visual cortex revealed by intrinsic signal imaging. *Proc Natl Acad Sci USA*, **93**: 9869–9874.